Targeting Trehalose and Methylglucose Lipopolysaccharide Biosynthetic Pathways in *M. tuberculosis*

Structural and Functional Characterisation, and Early-Stage Drug Discovery of OtsA and Rv3030

NUPUR VERMA

Lucy Cavendish College

September 2016

UNIVERSITY OF CAMBRIDGE

This Dissertation is submitted for the Degree of Doctor of Philosophy

Undertake difficult tasks by approaching what is easy in them.

Lao Tzu

DECLARATION

This dissertation is the result of my own work carried out in the Department of Biochemistry, University of Cambridge, between October 2012 and September 2016. It includes nothing that is the outcome of work done in collaboration except as declared in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University of similar institution. It does not exceed the prescribed limit of 60,000 words.

NUPUR VERMA

September 2016

ACKNOWLEDGEMENTS

I take this opportunity to express my profound sense of gratitude and respect to all those who helped me with my Ph.D. First and foremost, I offer my sincere respect and thanks to my supervisor, Prof. Sir Tom Blundell, for a giving me an opportunity to work with him, for his guidance and constant encouragement. He has been unhesitant in imparting his valuable wisdom, which assisted me in further developing my acumen for research. "*In loco parentis*", I still remember these kindest words he said to me when I just came to Cambridge. Yes, he truly has been one! His resolute support and motivation made these 4-years enjoyable to work, even during the dark phases.

I would like to thank Dr. Vitor Mendes and Dr. Michal Blaszczyk for their valuable help, guidance and supervision in the lab, for the stimulating scientific discussions, continous motivation and a great deal of learning experience. A big thanks to Dr. Sachin Surade, the former member of the lab, who considerably assisted me in settling down in Cambridge, imparted valuable scientific knowledge and whose advice kept me moving forward throughout. Dr. David Ascher warrants a special recognition; he has been an amazing mentor and friend, a person who had an unshakable faith in me and inspired me on various fronts, and for his constant willingness to discuss ideas. To Dr. Angela Pacitto for being a wonderful friend, and an impressively positive person whom I have always looked up to. I also thank the entire Blundell's group for their suggestions, team spirit, amiability, spreading the smiles around and making this journey delightful.

I am extremely grateful to my collaborator, Dr. Katherine Stott, for teaching me NMR spectroscopy and helping me in reviving the project, which seemed impossible otherwise. She has been graciously supportive and understanding. It has been an absolute pleasure to work with someone so brilliant and down to earth. Her vibrant personality was a confidence booster for sure.

My parents, across half the world, have been the best pillars. Thank you Mum and Dad for giving me the strength and encouraging me not to give up. Your tireless efforts, unconditional support and prudence have helped me sail through. To my younger brother, for his boundless patience in keeping me sane and his continual belief in me. Your presence, silent yet strong, and insights have strengthened me. Finally, to my friends, here in Cambridge and back home;

you have truly made every day bright, jolly and pleasurable! Last but not the least, thanks to the Almighty for bestowing upon me his choicest blessings.

ABSTRACT

Mycobacterium tuberculosis, the causative agent of tuberculosis in humans, is one of the most successful pathogens, with much of its virulence predominantly associated with its thick, lipid-rich cellular envelope that modulates the host immune response. Tuberculosis, having a long history, has proved to be a challenging disease to eradicate.

In mycobacteria, trehalose occurs as a free sugar in the cytoplasm, as well as a constituent of cell-wall glycolipids, such as cord-factor, sulfolipid-1 and lipooligosaccharides. *M. tuberculosis* possesses two pathways for synthesis of trehalose, the OtsA-OtsB pathway and the TreY-TreZ pathway. Multiple biosynthetic pathways underline the critical role that trehalose plays in the bacteria. OtsA-OtsB is the dominant pathway, required for bacterial growth in laboratory culture and for virulence in a mouse model. OtsA, a retaining glucosyltransferase encoded by the *otsA (Rv3490)* gene, catalyses the synthesis of trehalose-6-phosphate from ADP-glucose and glucose-6-phosphate, which is subsequently dephosphorylated by a functional homologue of OtsB (encoded by *otsB2/Rv3372* gene) to yield trehalose. Both *otsA* and *otsB2* gene have been shown to be essential for the growth of *M. tuberculosis*.

The present investigation focuses on this pathway, more specifically on OtsA. The recombinant orthologous protein from *M. thermoresistibile* was structurally and functionally characterised. The crystal structure revealed that there are two domains, with the catalytic site at their interface. The biochemical and structural data indicated that the enzyme had high preference for ADP-glucose as a glucose-donor. The protein activity was also shown to be regulated by feedback inhibition. Furthermore, a structure-guided, fragment-based drug-discovery exercise was carried out against the target, which led to identification of fragment hits that have an inhibitory effect on the activity of the enzyme.

Polymethylated polysaccharides (PMPSs) are complex intracellular polysaccharides, which are exclusive to mycobacteria and closely related species. PMPSs include methylglucose lipopolysaccharides (MGLPs) and methylmannose polysaccharides (MMPs). MGLPs, which are composed of acylated glucose and 6-O-methylglucose units, are found in all mycobacterial species, including the slow growing pathogenic *M. tuberculosis.* On the other hand, MMPs are linear polysaccharides that are not found in *M. tuberculosis*, but are found in

non-pathogenic *M. smegmatis* and other fast growing mycobacteria. Three gene clusters (*Rv1208-Rv1212c*, *Rv2418c-Rv2419c* and *Rv3030-Rv3037c*), initially thought to coordinate MGLPs biosynthesis, contain several genes that are considered to be essential for *M. tuberculosis* growth. One of the genes *Rv3030*, encoding for a SAM-dependent 6-O-methyltransferase, is essential for survival of the bacilli and plays a pivotal role in the biosynthesis of MGLPs.

In the present study, an orthologue of Rv3030 from *M. smegmatis* has been structurally characterised. To overcome a myriad of issues with crystallisation, low resolution and lack of reproducibility of the crystals, an alternative approach was taken to determine the structural features of the protein by NMR spectroscopy using ^{15}N, ^{13}C labelled protein. The protein has the β-sheet topology of the classic Rossmann fold, and exists as a monomer in solution.

OtsA and Rv3030, essential for survival of *M. tuberculosis*, represent attractive anti-tuberculosis drug targets, and this study focuses on understanding structural, biophysical and biochemical properties of these targets. This knowledge may then be used to identify lead chemical entities that bind to these proteins and modulate their function, providing new chemical tools that may be of use in designing antimicrobials for the fight against tuberculosis.

CONTENTS

LIST OF FIGURES

Chapter-1

Chapter-2

Chapter-3

Chapter-4

LIST OF TABLES

Chapter-1

Chapter-2

Chapter-3

Chapter-4

ABBREVIATIONS

3030M.smg	Rv3030 orthologue in *M. smegmatis*
3030M.thr	Rv3030 orthologue in *M. thermoresistible*
Å	angstrom
ADP	adenosine diphosphate
AG	arabinogalactan
AM	arabinomannan
APS	ammonium persulphate
ATCC	American Type Cell Culture
BLAST	basic local alignment search tool
bp	base pair
CD	circular dichroism
COSY	correlation spectroscopy
CPMG-NMR	Carr-Purcell-Meiboom-Gill nuclear magnetic resonance
CV	column volume
C_x	carbon chain containing x C-atoms
DGG	diglucosylglycerate
DLS	Diamond Light Source synchrotron (Oxford, UK)
DNA	deoxyribonucleic acid
DNAse-I	deoxyribonuclease I
dNTPs	dinucleotide triphosphates
DSMZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen
DTT	dithiothreitol
EDTA	ethylenediamine tetra acetic acid
eV	electron volt
FAS-I	fatty acid synthase complex-I
FBDD	fragment-based drug discovery
FID	free induction decay
FP	forward primer
GDP	guanidine diphosphate
GG	glucosylglycerate
GST	glutathione S-transferase

G6P	glucose-6-phosphate
HetNOE	heteronuclear NOE
His_6	6-histidine residues
HIV	human-immunodeficiency virus
HSQC	heteronuclear single quantum coherence spectroscopy
IC_{50}	half minimal inhibitory concentration
ITC	isothermal titration calorimetry
IMAC	immobilized metal affinity chromatography
IPTG	isopropyl-β-D-thiogalactopyranoside
kb	kilo base pair
kDa	kilo Dalton
K_m	Michaelis Menton constant
k_{cat}	turnover number of an enzyme
LAM	lipoarabinomannan
LB	Luria Bertani broth
LE	ligand efficiency
LM	lipomannan
LOS	lipooligosaccharides
M	meter
M	molar
MA	mycolic acids
MBP	maltose binding protein
min	minutes
MCS	multiple cloning site
mg	milligram
MGLP	methylglucose lipopolysaccharide
ml	milliliter
mm	milimeter
mM	millimolar
MMPs	methylmannose polysaccharides
MPD	2-Methyl-2,4-pentanediol
MRE	mean residue ellipticity
M. tb	*Mycobacterium tuberculosis*
MTs	methyltransferases

MW	molecular weight
MWCO	molecular weight cut-off
NADH	β-nicotinamide adenine dinucleotide (reduced)
NCBI	National Centre for Biotechnology Information
nm	nanomolar
ng	nanogram
Ni-NTA	nickle-nitrilo acetic acid
NDP	nucleotide diphosphate
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser effect spectroscopy
OD_{600}	optical density at 600 nm
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB	Protein Data Bank
PEG	polyethylene glycol
PG	peptidoglycan
pI	isoelectric point
PIMs	phosphatidylinositol mannosides
PMPSs	polymethylated polysaccharides
PMSF	phenylmethylsulfonyl fluoride
rcf	relative centrifugal force
RF	radio frequency
RNA	ribonucleic acid
RP	reverse primer
rpm	revolutions per minute
SAH	S-adenosyl homocysteine
SAM	S-adenosyl methionine
SAM-MTs	SAM-dependent methyltransferases
SAXS	small angle X-ray scattering
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate-polyacrylamide gel electrophoresis
sec	seconds
Se-met	seleno-methionine

SL-1	sulfolipid-1
S-SAD	sulphur single-wavelength anomalous diffraction
STD-NMR	saturation transfer difference nuclear magnetic resonance
SUMO	small ubiquitin-like modifier
SV-AUC	sedimentation velocity-analytical ultrcentrifugation
TB	tuberculosis
TEMED	N, N' tetraethyl methyl ethylene diamine
TDM	trehalose dimycolate
TMM	trehalose monomycolate
TOCSY	total correlation spectroscopy
TPS	trehalose-6-phospate synthase
TPP	trehalose-6-phospate phosphatase
TreS	trehalose synthase
TRIS	2-Amino-2-hydroxymethyl-propane-1,3-diol
TSA	thermal shift assay
T6P	trehalose-6-phosphate
UDP	uridine diphosphate
UV	ultraviolet
µl	microliter
µg	microgram
µm	micrometer
µM	micromolar

CHAPTER-1

General Introduction

1.1 The Genus *Mycobacterium* and Tuberculosis Infection

The genus *Mycobacterium* is one of the ancient bacterial genus, surmised to be originated 150 million years ago (Daniel, 2006). It includes 175 species (as of July 2016, List of Prokaryotic Names with Standing in Nomenclature, http://www.bacterio.net/), most of which are saprophytes without clinical relevance, but there are number of opportunistic and strict pathogens that infect humans and animals (Rastogi *et al*., 2001).

Mycobacterial infections trace back to antiquity, impacting humankind and advancements in health care research. World's deadly diseases, tuberculosis and leprosy, caused by *M. tuberculosis* (*M. tb*) and *M. leprae* respectively, have plagued humankind for centuries (Palomino *et al*., 2007). The tubercule bacilli was discovered by Robert Koch on 24th March 1882, refuting the idea that the disease tuberculosis (TB) was hereditarily acquired or was only affecting people of low-social strata living in unhealthy conditions (Hass and Hass, 1996).

TB, earlier also known as the *white death* and the *great white plague*, was epidemic in Europe and America during 18th and 19th centuries, and by the end of 19th century, there was also a rise in incidence in India and China (Palomino *et al*., 2007). Though improvement in general public health, and advent of early diagnostic techniques and anti-tuberculosis drugs, reduced the burden, the disease was never eradicated, and around 1985, there was upsurge in number of TB cases in industrialised countries (Frith, 2014). There has also been emergence of infection by opportunistic non-tuberculous mycobacteria inhabiting natural reservoirs, which is now being recognized worldwide (Cook, 2010; Falkinham, 2009).

Mycobacteria are acid-fast, obligatory aerobic actinomycetes, placed within the CNM (Corynebacterium-Nocardia-Mycobacterium) complex on the basis of chemotaxonomic studies (Lechevalier and Lechevalier, 1970). Mycobacteria form straight or slightly curved, non-motile and non-sporulated rods that grow as filaments, have unusually high content of guanine and cytosine in its genomic DNA, and a unique lipid rich cell wall composed mainly of mycolic acids, that differentiates it from other closely related genera (Palomino *et al*., 2007). The species within this genus exhibit pronounced diversity, inhabiting diverse biotopes with only a handful of them causing pathogenicity. Slow growing species, like *M. tuberculosis* and *M. leprae*, are obligatory intracellular pathogens triggering infection in

humans and animals; whereas rapid growers, like *M. smegmatis*, are environmental saprophytes and typically non-pathogenic (Palomino *et al*., 2007). Several species form white or cream coloured colonies, but many fast growing species also form orange or bright yellow colonies containing carotenoid pigments (Zureck, 1985). In most species, the pigments are formed only in dark (scotochromogenic species), but some species form pigments only in response to light (photochromogenic species) (Dworkin *et al*., 2006).

M. tb poses a great threat to humanity and harbour in one-third of global society, either as the carrier of latent infection or as the active disease (Dye, 2009). Tuberculosis is one of the leading causes of mortality worldwide due to bacterial infection, and it is estimated that in 2014, there were 9.6 million new TB cases and 1.5 million deaths, out of which 0.4 million were among HIV-positive patients (Global tuberculosis report, WHO, 2015). The disease manifests itself primarily in lungs, though cases of extra-pulmonary TB are not uncommon (Kaufmann, 2001). The pathogen has evolved mechanistically and modulates expression of various metabolism-related genes to inhabit alveolar macrophage, wherein all other microbes perish due to microbicidal activity (Hingley-Wilson *et al*., 2003; Kaufmann, 2001).

Despite considerable progress in the health care and medical sector, the rapid emergence in drug resistant strains of *M. tb*, and co-epidemics of TB and HIV has exacerbated the effect of TB, making it one of the most challenging diseases to control and eradicate (Global tuberculosis report, WHO, 2015; Wallis *et al*., 2016; Zumla *et al*., 2012). Thus, there is an urgent need to develop novel therapeutics for treatment of this insidious disease, which requires identification of novel pathways and new drug targets.

1.2 Mycobacterial Cellular Envelope

Mycobacteria has a unique, resilient cell envelope having three distinct layers: an outer coat or capsule, a cell wall and an inner plasma membrane **(Figure 1.1)** (Brennan, 2003). This ensemble of free and bound polysaccharides, lipids, proteins and their conjugates confers much of the pathogenicity and drug-resistance to mycobacteria, and also modulates the host immune response (Hett and Rubin, 2008; Riley, 2006). The outer capsule, and the cell wall rich in long chain fatty acids known as mycolic acids (MA), are unique elements of mycobacteria that differentiates it from other prokaryotes (Dworkin *et al*., 2006). The composition of MA in the cell wall further individualize mycobacteria from closely related, MA containing genera *Corynebacterium* and *Nocardia* (Dworkin *et al*., 2006).

Deciphering of the *M. tb* genome along with advancements in analytical techniques has helped in better understanding of the basic genetics, biosynthesis and structure of mycobacterial cell envelope (Brennan, 2003).

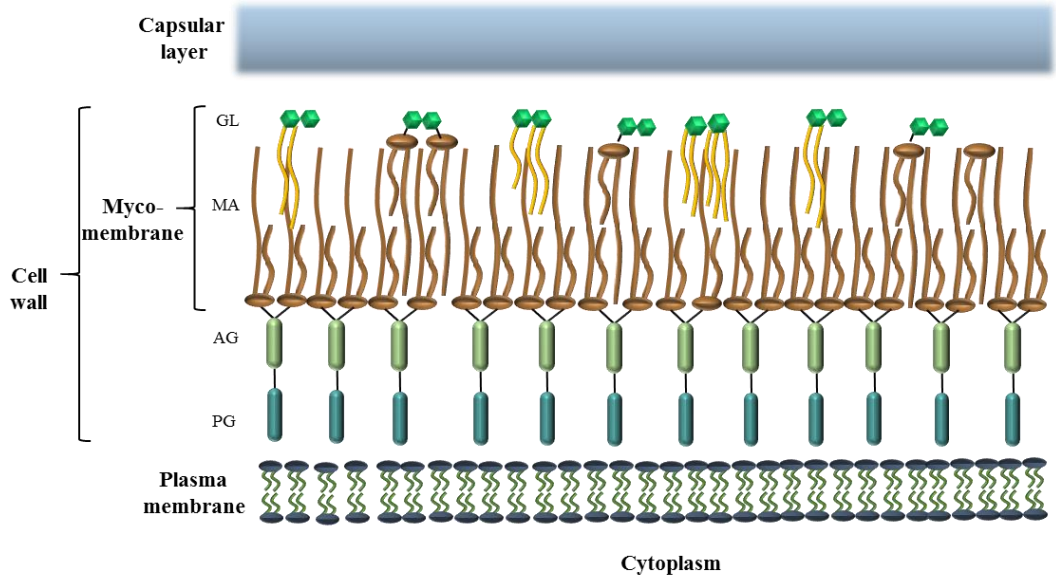


FIGURE 1.1 The schematic representation mycobacterial cell envelope.
GL: glycolipids, MA: mycolic acid, AG: arabinogalactan, PG: peptidoglycan (Figure adapted from Maranha *et al*., 2015; Nobre *et al*., 2014).

1.2.1 The Capsule

The surface of mycobacteria is a loosely bound structure known as capsule, which differs in composition between the pathogenic and the non-pathogenic species of mycobacteria (Hett and Rubin, 2008; Sani *et al*., 2010). The previously debated existence of this layer was confirmed when a thick capsule produced by pathogenic mycobacteria was detected *in vitro* by plunge freezing cryo-electron microscopy technique (Ortalo-Magne *et al*., 1995; Sani *et al*., 2010; Zuber *et al*., 2008). 97% of capsular matter is composed of polysaccharides and proteins, containing miniscule amounts of lipids (PIMs and LAM) as well (Sani *et al*., 2010).

Three types of polysaccharides found in the capsule of *M. tb* are: arabinomannan (AM) consisting of a 6-α-D-Man-l mannan core and arabinan termini capped with mannosyl residues; a D-mannan composed of α-(1→6)-linked D-mannose units; and a high molecular weight α-D-glucan, which is analogous to glycogen (Lemassu and Daffé, 1994; Ortalo-Magne *et al*., 1995). The high molecular weight (>100 kDa) α-D-glucan, which is the chief constituent of this labile layer amounting to 90% of the polysaccharides, is composed of α-(1→4)-linked D-glucose units, branched every five or six residues by oligoglucosides (Angala *et al*., 2014; Lemassu and Daffé, 1994). Some of the enzymes involved in biosynthesis of glucan and glycogen, are also involved or have putative function in synthesis of methylglucose lipopolysaccharides (MGLPs), an intracellular polysaccharide (described in detail in section 1.4), thus suggesting that the capsular α-D-glucan is synthesized in the cytoplasm and then transported to the surface by unknown carriers (Kalscheuer *et al*., 2010; Kaur *et al*., 2009a; Stadthagen *et al*., 2007).

These capsular polysaccharides have been implicated in several biological roles, including host-pathogen interactions, immunomodulation and immunopathogenesis (Angala *et al*., 2014; Kaur *et al*., 2009a). α-D-glucan has also been propounded to be involved in bacilli evasion of the immune system by molecular mimicry (Lemassu and Daffé, 1994). α-D-glucan mediates nonopsonic binding of *M. tb* to complement receptor-3 (CR-3), one of the principal phagocytic receptors of monocytes and neutrophils and this CR3-mediated phagocytosis promotes intracellular survival by suppressing IL-12 production and diminution of respiratory burst (Cywes *et al*., 1997; Ehlers and Daffé, 1998). However, since most of these studies have been elucidated *in vitro*, pivoting on the purified capsular glucan with cellular models,

α-D-glucan may not be accurately pertinent in mycobacterial infections, and warrants further investigation to define its role in *M. tb* pathogenicity and survival (Kaur *et al*., 2009a).

1.2.2 The Cell Wall

Mycobacterial cell wall is a complex, bi-layer structure that forms a stringent permeability barrier (Hett and Rubin, 2008). The insoluble core of the cell wall, the mycolyl-arabinogalactan-peptidoglycan (mAGP) complex, consists of covalently linked peptidoglycan (PG), arabinogalactan (AG) and mycolic acids, sequentially extending outwards from the plasma membrane **(Figure 1.2)** (Brennan, 2003; Hett and Rubin, 2008). Also present therein are free soluble-lipids, cell-wall proteins, the phosphatidylinositol mannoside (PIM), the phthiocerol-containing lipids, lipomannan (LM) and lipoarabinomannan (LAM) (Riley, 2006). The mycobacterial cell wall has received special attention from the perspective of drug-development and diagnostics, with several anti-tubercular agents targeting the biosynthesis of ingredients of the cellular envelope.

Peptidoglycan

Peptidoglycan (PG) or murein, is an intricate glycopolymer forming a rigid coat outside the plasma membrane that allows the bacteria to maintain its shape, resist the effects of osmotic pressure and serves as a foundation for other cell wall structures (Kaur *et al*., 2009a). The glycan backbone is a nexus of alternating units of N-acetylglucosamine and modified muramic acid in a linear β-(1,4) linkage, with tetra-peptide side chains that are cross-linked, providing auxiliary structural cohesion to the cells (Brennan, 2003). There has been a renewed interest in PG synthesis as a promising target for novel chemotherapeutics (Angala *et al*., 2014).

Arabinogalactan

Arabinogalactan (AG) is the major cell-wall polysaccharide of mycobacteria, attached to PG by a phosphoryl-N-acetylglucosaminosyl-rhamnosyl linkage (Bhamidi *et al*., 2008; Brennan, 2003). It is vital for cell wall integrity and for tethering mycolic acid chains to PG layer (Brennan, 2003). AG is an assembly of arabinan and galactan, both in furanose form, with approximately 30 galactose residues arranged in an alternating (1→5), (1→6) linkages, and arabinan is attached to position 5 of the galactan in (1→5) linkages (Hett and Rubin, 2008). The anti-microbial action of ethambutol, a first-line anti-TB drug that inhibits synthesis of arabinan domains of LAM and AG, demonstrates the importance of AG in *M. tb* (Belanger *et al*., 1996; Goude *et al*., 2009).

Mycolic Acids

The outer layer of the mycobacterial cell wall, also referred as the mycomembrane, is concocted of C_{60}-C_{90} α-branched, β-hydroxylated mycolic fatty acids that are covalently linked to the arabinogalactan-peptidoglycan layer; and free glycolipids intercalating with mycolic acids (MA) (Sani *et al*., 2010). This highly insoluble network of fatty acids is the predominant feature that contributes to low permeability of the cell wall and to the intrinsic resistance to drugs (Brennan, 2003; Hett and Rubin, 2008). Essentiality of MA in survival and virulence of the bacilli is reflected by the complexity and variability of MA biosynthetic pathway, and by the fact that isoniazid, the prime first-line antituberculous drug, targets MA biosynthesis (Cantaloube *et al*., 2011). Intracellular MGLPs form stable 1:1 ratio complexes with fatty acyl chains and acyl-CoAs, and regulate the activity of mega-enzyme FAS-I, involved in biosynthesis of MA, thus deriving the rationale that fatty acid synthesis as well is controlled by the intracellular polysaccharides (Flick and Bloch, 1974; Mendes *et al*., 2012).

Glycoconjugates of Cell Wall

The mycobacterial cell wall contains a high content of lipoglycans and lipoglycoproteins, phosphatidyl-myo-inositol mannosides (PIMs), lipomannan (LM) and lipoarabinomannan

(LAM) (Kaur *et al*., 2009a). These ubiquitous molecules amalgamate non-covalently with both the inner and outer layers of cell envelope, through their phosphatidyl-myo-inositol (PI) moiety (Ortalo-Magne *et al*., 1995; Pitarque *et al*., 2008). The PI anchor is conserved in these lipoglycans, suggestive of their metabolic relationship, and is heterogenous with variations occuring in number, location and nature of fatty acids (Besra and Brennan, 1997; Kaur *et al*., 2009a). PIMs may contain one to six mannoses attached to inositol, but the major ones in *M. tb* are PI-dimannosides (Ac_1PIM_2 and Ac_2PIM_2) and PI-hexamannosides (Ac_1PIM_6 and Ac_2PIM_6) (Gilleron *et al*., 2001, 2003). Structurally, LMs and LAMs are extensions of PIMs. In addition to the PI moiety, LAM and LM possess a common linear α-(1,6)-linked mannan backbone made of 20-25 Man*p* residues, occasionally branched at C-2 by single Man*p* units; and in case of LAM, there is an additional arabinan moeity (Chatterjee and Khoo, 1998; Chatterjee *et al*., 1992; Gilleron *et al*., 2000). These glycoconjugates are involved in promoting entry of *M. tb* inside phagocytes, modulating the immune response, apoptosis and regulating phagosome maturation *in vitro*, but because these studies have been done with purified molecules and cellular models, it may not reflect their true significance (Appelmelk *et al*., 2008; Briken *et al*., 2004; Kaur *et al*., 2009a; Torrelles *et al*., 2008). PIMs regulates cell division, act as permeablity barrier and maintains the integrity of cell wall (Parish *et al*., 1997; Patterson *et al*., 2003).

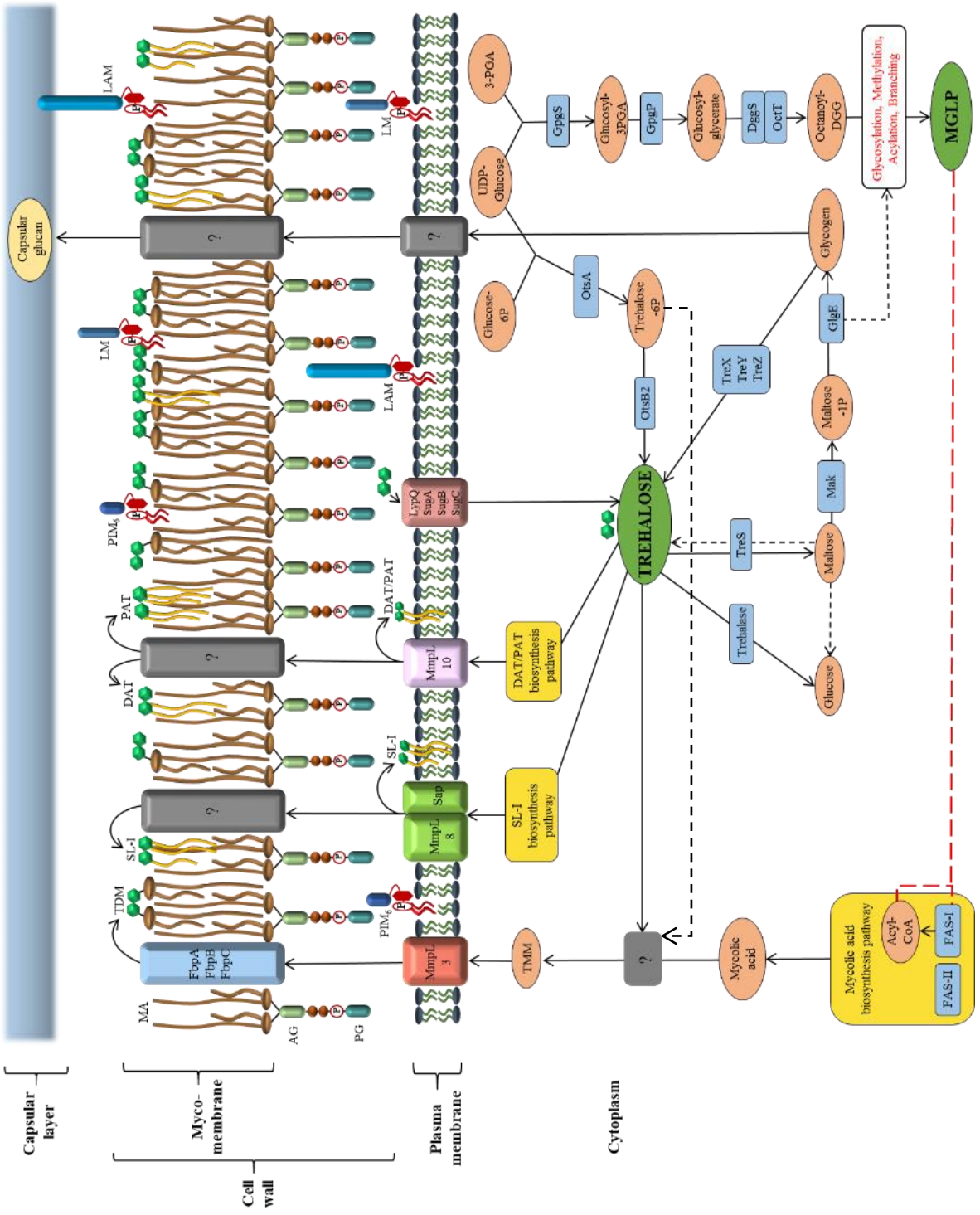


FIGURE 1.2 The mycobacterial cell envelope.

Detailed representation of capsule, cell wall and cytoplasmic section. Representation includes pathways and enzymes involved in biosynthesis of trehalose, trehalose-conjugates of cell wall, and MGLPs.
MA: mycolic acid,
TDM: trehalose dimycolate,
TMM: trehalose monomycolate,
SL-1: sulfolipid-1,
DAT: diacyltrehalose,
PAT: pentaacyltrehalose,
PIM: phosphoinositol mannoside,
LAM: lipoarabinomannan,
LM: lipomannan,
AG: arabinogalactan,
PG: peptidoglycan,
Acyl-CoA: acyl co-enzyme A,
FAS: fatty acid synthase,
P: phosphate,
3-PGA: D-3-phosphoglyceric acid,
DGG: diglucosylglycerate.
Enzymes are shaded blue. Dashed arrows represent putative function. Red dashed line represents the complex formation between acyl-CoA and MGLP, and modulation of FAS-I activity (see section 1.4.3 for details). (Figure adapted from Crellin *et al.*, 2013; Maranha *et al.*, 2015; Nobre *et al.*, 2014).

1.3 Trehalose: From Mycobacterial Perspective

Trehalose (α-D-glucopyranosyl-(1,1)-α-D-glucopyranoside) is a non-reducing disaccharide, abundantly found in plants, insects, yeast, bacteria, fungi and invertebrates (Becker *et al*., 1996; Gancedo and Flores, 2004; Grennan, 2007; Strom and Kaasen, 1993; Wright and Marshall, 1971). It exhibits unusual stability because of the two linked glucopyranose rings occurring at the reducing end of the glycosyl residues (α-carbons), and its glycosidic bond is not cleaved by α-glycosidase (Ohtake and Wang, 2011).

Trehalose carries out divergent roles in different species, and its production is commonly related to physiological stresses. The most prevalent effect is its potential to allow resurrection of plants from complete desiccation due to its unique water-retention property (Ohtake and Wang, 2011). Trehalose is also produced in response to osmotic stress and is critical for thermal tolerance in many organisms (Hottiger *et al*., 1987). In insects, trehalose is the major source of energy, and is involved in anoxic tolerance and development (Chen *et al*., 2002). Apart from being the energy source and an osmoprotectant in several bacteria as well, trehalose has noted structural and functional importance in *Mycobacterium tuberculosis*. In mycobacteria, trehalose is found abundantly in cytosol and cell wall, both as a free disaccharide and as glycoconjugate, constituting approximately 1.5-3% of total dry weight of cell in *M. smegmatis* (Elbein, 2003).

1.3.1 Trehalose Metabolism in Mycobacteria

It has been widely contemplated that mycobacteria has three pathways for trehalose synthesis: OtsAB, TreYZ and TreS (Avonce *et al*., 2006; Elbein, 2003). Though mycobacteria can principally exploit all the three pathways, OtsAB is the dominant pathway for trehalose production in *M. tb* (De Smet *et al*., 2000; Murphy *et al*., 2005). Recent studies in *M. smegmatis* have shown that trehalose synthase (TreS) is not the trehalose producing pathway, and its overexpression in *M. tb* decreases the level of trehalose (Miah *et al*., 2013). The multiple trehalose biosynthetic pathways underline its crucial role in the bacteria.

OtsA-OtsB Pathway

This is the most dominant pathway that catalyses production of trehalose from NDP-glucose and glucose-6-phosphate by a two-step process. In the first step, OtsA enzyme (trehalose-6-phosphate synthase, TPS) catalyses the condensation reaction between NDP-glucose and glucose-6-phosphate to produce an intermediate, trehalose-6-phosphate, and this phosphorylated compound is then dephosphorylated by OtsB enzyme (trehalose-6-phosphate phosphatase, TPP) into trehalose **(Figure 1.3)** (De Smet *et al*., 2000).

In *M. tuberculosis* H37Rv, genes *otsA* and *otsB2* have been demonstrated to be essential for growth by saturation transposon mutagenesis, and are required for survival (Griffin *et al*., 2011; Sassetti *et al*., 2003). Trehalose-6-phosphate synthase (OtsA) is encoded by the *otsA* gene (*Rv3490*), and its mutants showed severe growth defects *in vitro* and *in vivo* (Murphy *et al*., 2005). Furthermore, genome analysis of *M. tb* has identified two open reading frames of the *otsB* gene: *otsB1* (*Rv2006*) and *otsB2* (*Rv3372*), which encode putative trehalose-6-phosphate phosphatase (Cole *et al*., 1998; Murphy *et al*., 2005). *otsB1* is part of a regulon, which is induced by exposure to hypoxia or to low concentrations of nitric oxide, its product has no detectable phosphatase activity, and is associated with non-replicating survival of the bacteria (Edavana *et al*., 2004; Voskuil *et al*., 2003). *otsB2*, the functional homologue, has been reported to be strictly essential for growth of the bacilli, and mutation of *otsB2* gene caused lethality and validated that phosphatase activity of OtsB is carried out only by this homologue, and the supply of exogenous trehalose did not reverse the phenotype (Murphy *et al*., 2005). Moreover, OtsB2 was also shown to induce humoral and cellular immune responses, and may be a candidate for vaccine development for control of TB disease (Zhang *et al*., 2007).

TreY-TreZ Pathway

This pathway, first identified in *Rhizobia*, *Arthrobacter* and *Sulfolobus acidocaldarius*, produces trehalose from terminal maltoses in glycogen chains by the activity of products of *treY* and *treZ* genes (Maruta *et al*., 1996c, 1996a, 1996b). TreY, a maltooligosyltrehalose trehalohydrolase encoded by *treY* gene, catalyses the rearrangement of the α-(1→4)-linkage

of the terminal disaccharide at the reducing end in glycogen into an α-(1→1)-linkage, and then the trehalose disaccharide is subsequently hydrolysed by the product of *treZ* gene, maltooligosaccharyltrehalose trehalohydrolase **(Figure 1.3)** (Woodruff *et al*., 2004).

The gene cluster in *Mycobacterium tuberculosis* has two genes *Rv1563c* (*treY/ glgY*) and *Rv1562c* (*treZ/glgZ*) adjoining, with an additional gene *treX/ glgX* (*Rv1564c*) encoding a putative glycogen-debranching enzyme, though the function of the latter still remains to be validated (Cole *et al*., 1998; De Smet *et al*., 2000; Seibold and Eikmanns, 2007). Though this pathway is predominant in *C. glutamicum* (Tzvetkov *et al*., 2003), genetic inactivation of *treY/treZ* genes has no effect on *M. tb* growth *in vitro* (Murphy *et al*., 2005). In *M. leprae*, *treY* is a pseudogene, leaving OtsAB the only intact pathway for trehalose production (Cole *et al*., 2001).

TreS Pathway

Trehalose production by this pathway initially involved catalysis by the enzyme trehalose synthase (TreS), which isomerized the α-(1→4)-linkage of maltose into an α-(1→1)-linkage in trehalose, using free maltose as a substrate, unlike TreY (Elbein, 2003). Though it was considered one of the pathways for trehalose biosynthesis in mycobacteria, recent genetic studies have confuted the role of TreS in synthesis of trehalose from maltose *in vivo*, and ^{1}H- and ^{19}F-NMR analyses confirm that flux through TreS flows from trehalose to α-maltose, and overexpression of TreS led to a decrease in trehalose levels (Miah *et al*., 2013). While knock-out mutants of *treS* (*Rv0126*) in *M. tb* showed no evident growth impairment compared to wild type during initial phase, they did lead to noticeable reduction in the bacterial load during late and near time-to death stages of infections in mice, thus indicating TreS could have a role only in the late-stage pathogenesis of the disease (Murphy *et al*., 2005).

Another function of TreS that has been suggested is the α-amylase activity, wherein the enzyme was proposed to catalyse hydrolysis of glycogen with an α-amylase domain (Pan *et al*., 2008). But this was disproved when it was shown that the expected active site residues were absent from the acarbose binding site, and this activity in mycobacteria was only observed *in vitro* and at a very low rate, suggestive of a non-physiological role (Caner *et al*., 2013). Also, TreS has two different active sites that would explain the lack of α-amylase

activity without affecting the observed trehalose-maltose interconverting activity (Pan *et al*., 2004).

Trehalose Catabolism

M. tb possesses a trehalase-encoding gene (*Rv2402*), in addition to TreS, that directly hydrolyses trehalose, to balance the intracellular levels of trehalose during specific growth conditions (Elbein *et al*., 2010; Pan *et al*., 2004). Though the importance of trehalase activity and hypothetical regulation with TreS-Mak-GlgE pathway (described in section 1.5) *in vivo* remains elusive, these mechanisms can be significant for rapid assemblage of carbon and energy reserves in response to stress, or during osmoadaptation, when tight regulation of trehalose levels is required (Carroll *et al*., 2007; Elbein *et al*., 2010). In *M. tb*, the trehalase-encoding gene was not considered essential by saturation transposon mutagenesis which hypothetically posits an overlapping function with TreS (Griffin *et al*., 2011). But contrastingly, a trehalase-negative mutant *M. smegmatis* could not be retrieved, advocating its essentiality for growth (Carroll *et al*., 2007).

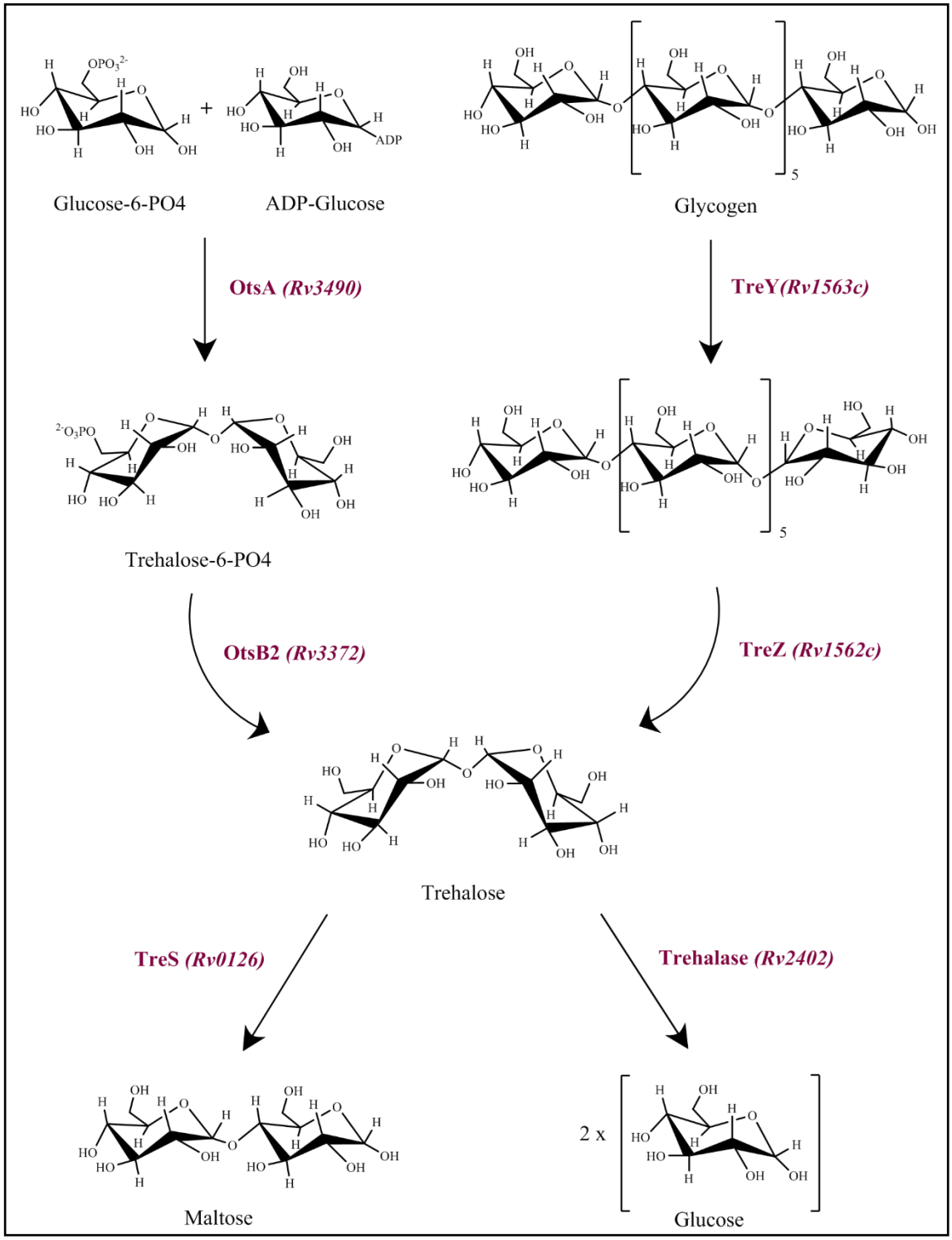


FIGURE 1.3 Trehalose metabolism in *M. tuberculosis*.

Enzymes are indicated in purple. *OtsA*: trehalose 6-phosphate synthase; *OtsB2*: trehalose 6-phosphate phosphatase; *TreY*: maltooligosyl-trehalose synthase; *TreZ*: maltooligosyl-trehalose trehalohydrolase. (Figure adapted from Carroll *et al*., 2007; Miah *et al*., 2013; Woodruff *et al*., 2004).

1.3.2 Biological Importance of Trehalose in Mycobacteria

Mycobacterium tuberculosis produces trehalose for an array of tasks, through various independent pathways utilising different substrates. In spite of that, OtsAB has been validated to be the prime route for generating trehalose, and alternate pathways cannot compensate its depletion. Trehalose is a constituent of the mycobacterial cytoplasm and trehalose-containing compounds are abundant in mycobacterial cell wall, many of which are species-specific (Gautier *et al*., 1992; Jackson *et al*., 2007; Nobre *et al*., 2014).

Trehalose Mycolates

Trehalose-containing glycolipids in the cell envelope either protect mycobacteria from harsh environmental conditions or act as important virulence factors during pathogen-host interactions (Etienne *et al*., 2009; Harland *et al*., 2008; Stanley and Cox, 2013; Welsh *et al*., 2013). The most prevalent glycolipids are trehalose-6-mycolate (trehalose monomycolate/ TMM) and trehalose 6, 6'-dimycolate (TDM) **(Figure 1.4 A and B)**, which play a pivotal role in the low permeability and high resistance of mycobacteria to many antibiotics (Ortalo-Magne *et al*., 1995). TMM is a ubiquitous metabolite in mycobacteria, implicated as a carrier of mycolic acid from the site of synthesis in the cytosol to the acceptor arabinogalactan in the cell wall (Dhariwal *et al*., 1987).

TDM intercalates with the AG-linked MAs in the outer membrane, and confers to the envelope integrity and mediates interactions with the host's immune system (Nobre *et al*., 2014; Welsh *et al*., 2013). A potent antigenic "cord factor", derived from trehalose dimycolate by modification of its lipid chains, has been associated with virulence of pathogenic mycobacteria (S. Glickman, 2008). TDM modifies macrophage gene expression, prevents phagosome-lysosome fusion and inhibits phagolysosome acidification that favours bacterial survival inside macrophages (Sakamoto *et al*., 2013). It is also the sole mycobacterial constituent that has potent immune-stimulatory activity, which imparts granuloma formation, adjuvant activity and tumour regression properties (Lang, 2013; Welsh *et al*., 2013). Recent investigations on the dynamic properties of TDM as a function of

temperature indicated that higher rigidity of trehalose mycolates and interaction between mycolates was related to the trehalose molecule itself (Migliardo *et al*., 2015).

Acyltrehaloses

M. tb produces a multitude of other different surface-exposed acyltrehaloses, namely the sulfatides, diacyltrehaloses (DATs), triacyltrehaloses (TATs) and pentacyltrehaloses (PATs) **(Figure 1.4 C and D)** (Daffé *et al*., 1988; Lemassu *et al*., 1991; Muñoz *et al*., 1997), which are complex glycolipids made of trehalose esterified with up to five multiple methyl-branched long-chain fatty acids (Rousseau *et al*., 2003). Mycolipenic acids, the major acyl substituents found in PAT, TAT and some forms of DAT, have been found in the pathogenic *M. tb*, but not in the avirulent strain *M. tuberculosis* H37Ra or in the attenuated vaccine strain *M. bovis* BCG (Cason *et al*., 1956). This is indicative that these moieties play a crucial role in virulence of *M. tb*. These trehalose esters are potent inhibitors of leucocyte migration *in vitro* and DAT inhibits the proliferation of murine T-cells (Saavedra *et al*., 2001). Moreover, due to antigenicity of these fatty acids, they are candidate antigens for the serodiagnosis of TB (Besra *et al*., 1992; Lemassu *et al*., 1991). Though studies with purified molecules in cellular models implicate a significant role of acyltrehaloses in virulence of *M. tb*, their relevance during the natural process of infection remains to be unravelled (Jackson *et al*., 2007; Rousseau *et al*., 2003).

Sulfolipid-1

The outer membrane of *M. tb* abundantly contains sulfolipid-1 (SL-1), which is a tetraacylated trehalose-2-sulfate glycolipid **(Figure 1.4 E)**, and the 50-year long search for its biological relevance has attributed in myriad, occasionally conflicting, roles in mycobacteria (Goren, 1972; Rousseau *et al*., 2003). It has been suggested that in cell-culture models, purified SL-1 activates and suppresses the production of cytokines and reactive oxygen species produced by human leukocytes (Pabst *et al*., 1988; Zhang *et al*., 1988). SL-1 are lysosmotropic and alter phagosome-lysosome fusion, thereby allowing the bacilli to survive inside macrophages (Goren *et al*., 1976). The levels of SL-1 are directly correlated with

virulence, with this sulfated-glycolipid being unique to the *M. tb* complex (Gangadharam *et al*., 1963). SL-1 in concert with cord factor disrupt mitochondrial oxidative phosphorylation, and both trehalose containing glycolipids damage mitochondrial structure and function, forming basis of the synergistic toxicity *in vivo* (Kato and Goren, 1974). Anionic SL-1 also regulates the intracellular growth of bacteria in a species-specific fashion, and arbitrates its susceptibility to a human cationic antimicrobial peptide LL-37 *in vitro* (Gilmore *et al*., 2012; Nobre *et al*., 2014).

Lipooligosaccharides

Mycobacterial lipooligosaccharides (LOS) are a distinct class of trehalose-containing glycolipids **(Figure 1.4 F)**, composed of oligosaccharide chains linked to trehalose, and acylated with polymethylated-branched fatty acid (Etienne *et al*., 2009). LOS, found in many mycobacterial species such as *M. kansasii*, the Canettii strain of *M. tuberculosis*, *M. marinum, M. gastri,* display substantial structural variations in the lipid moiety and the glycan core (Daffe *et al*., 1991; Hunter *et al*., 1983, 1988). Extensively reported structures of LOS from *M. marinum* reveals that they encompass a string of subspecies, LOS-I to LOS-IV, which are sequential intermediates that differ in number of monosaccharide units (Rombouts *et al*., 2009, 2010, 2011). These highly antigenic cell-surface exposed glycolipids have been implicated in infection of host macrophages, sliding motility and biofilm formation (Ren *et al*., 2007). LOS are liable to interfere with inception of proinflammatory response, by inhibiting the secretion of TNF-ɣ in LPS-stimulated human macrophages *in vitro* (Alibaud *et al*., 2014).

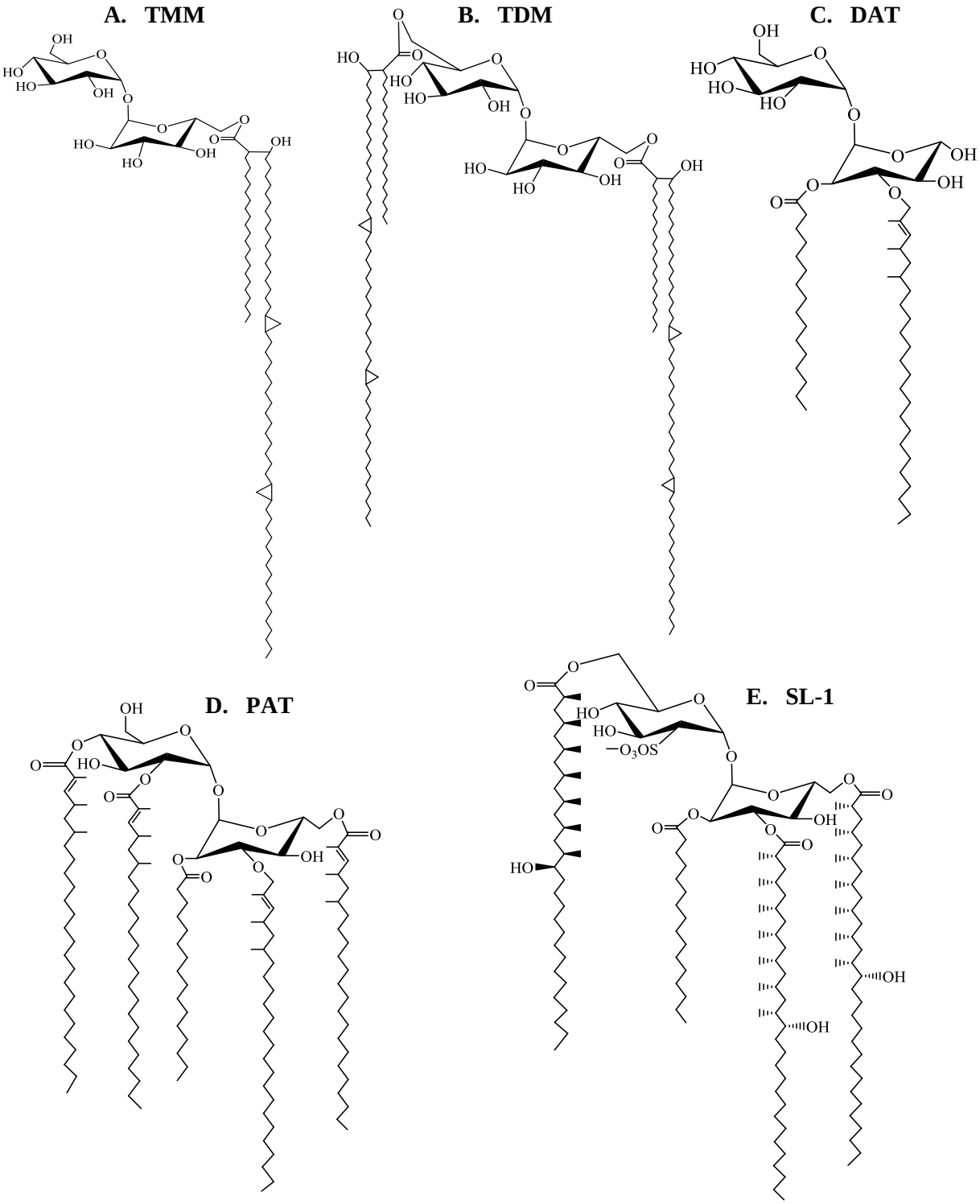


FIGURE 1.4 Structures of trehalose-conjugates in mycobacteria.

A. TMM: trehalose monomycolate
B. TDM: trehalose dimycolate
C. PAT: pentaacyltrehalose
D. DAT: diacyltrehalose
E. SL-1: sulfolipid-1
F. LOS: lipooigosaccharide (tetrasccharide core of LOS from *M. kansasii*)
(Figure adapted from Nobre *et al*., 2014; Woodruff *et al*., 2004).

1.4 Methylglucose Lipopolysaccharides

Mycobacteria produce complex polymethylated polysaccharides (PMPSs) (Berg *et al*., 2007), which were initially isolated in Clinton Ballou's laboratory from *Mycobacterium phlei* (Gray and Ballou, 1971). These are cytoplasmic lipo-polysaccharides, composed of 10-20 sugar units, many of which are partially *O*-methylated (Jackson and Brennan, 2009). Two types of PMPSs found in mycobacteria and closely related species are 3-O-methylmannose polysaccharides (MMPs) and 6-O-methylglucose lipopolysaccharides (MGLPs) (Jackson and Brennan, 2009; Mendes *et al*., 2012).

MMPs are composed of 10-13 units of 3-O-methyl-D-mannoses linked via α-(1→4)-linkages, with a single α-(1→4)-linked unmethylated D-mannose present at its non-reducing and an α-methyl aglycon at the reducing end **(Figure 1.5)** (Mendes *et al*., 2012). These are found in non-pathogenic, fast-growing species of mycobacteria (Weisman and Ballou, 1984) and *Streptomyces griseus*, though the latter have acetylated molecules (Harris and Gray, 1977). MMPs however, have not been observed in *M. tb* and this indicates that they are probably restricted to the fast-growing mycobacterial species, unlike MGLPs (Stadthagen *et al*., 2007).

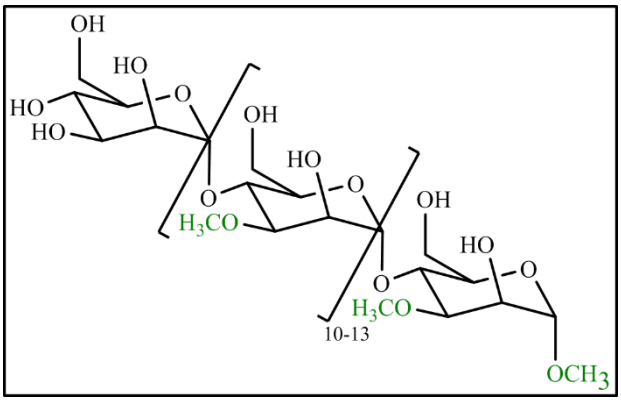


FIGURE 1.5 Structure of 3-O methylmannose polysaccharides (MMPs) from *M. smegmatis*. Methyl groups are highlighted in green (Figure adapted from Mendes *et al.*, 2012).

MGLPs are polysaccharides of glucose and methylglucose that are esterified with short-chain fatty acids (Ferguson and Ballou, 1970). These are a heterogeneous mix of closely related polysaccharides, differing in their degree of glycosylation, number of methylation sites, and on the nature and frequency of the esterified acyl chains (Tuffal *et al*., 1995). MGLPs have been identified in a few *Nocardia* species and in *M. tuberculosis*, *M. bovis* BCG, *M. leprae*, *M. smegmatis* and *M. xenopi*, and the slow-growing, pathogenic mycobacteria may contain only this type of PMPSs (Gray and Ballou, 1971; Jackson and Brennan, 2009; Mendes *et al*., 2012).

1.4.1 Composition and Structure of MGLPs

The sugar constituents of MGLPs are D-glucose, 6-O-methyl-D-glucose, 3-O-methyl-D-glucose and D-glyceric acid; and the lipid components are propionate, isobutyrate, succinate, and octanoate (Lee, 1966; Lee and Ballou, 1964).

The main MGLP chain contains an assortment of α-(1→4)-linked glucose units and ten 6-O-methylglucose units, with two branching glucoses linked via β-(1→3)-linkage to the glucoses 1 and 3 of the MGLPs main chain (Mendes *et al*., 2012). A 3-O-methylglucose is present at the non-reducing end, whereas, the reducing end is composed of α-D-glucopyranosyl-(1-2)-D-glyceric acid (glucosylglycerate, GG), which is linked to the first glucose of the main MGLP chain via α-(1→6) linkages and form diglucosylglycerate (DGG) **(Figure 1.6)** (Mendes *et al*., 2012).

Acetate, propionate and isobutyrate are esterified to the methylglucose units close to the non-reducing end; and succinate to both the branching glucoses (Keller and Ballou, 1968; Smith and Ballou, 1973). Early results by Smith and Ballou showed that in preparations of MGLPs from *M. phelei*, an octanoate moiety was attached to the primary hydroxyl group of the glycerate at the reducing end (Smith and Ballou, 1973). However, a recent study has redefined the architecture of this end, according to which octanoate is present at the C6-OH position of the glucoses of GG and DGG, and not to the glycerate molecule **(Figure 1.6)** (Maranha *et al*., 2015).

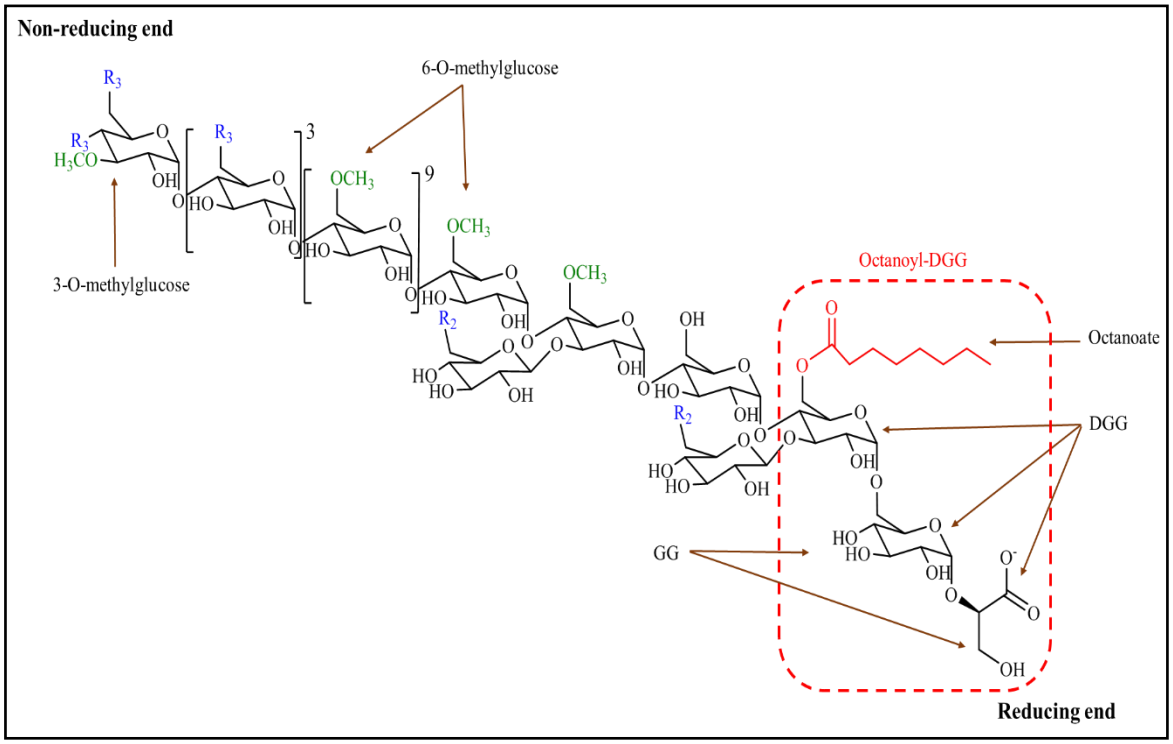


FIGURE 1.6 Structure of methylglucose lipopolysaccharides (MGLPs).

Octanoyl-DGG highlighted by red dashed box, represent new architecture of the reducing end of the MGLP chain, based on the recent findings (described in the text). R2 and R3 are acyl groups (blue). R_2: succinate; R_3: acetate, propionate or isobutyrate. Methyl groups are highlighted green. (Figure adapted from Maranha *et al*., 2015; Mendes *et al*., 2012).

In solution, MGLPs conform to a helical structure because of the stereotypical constraints conferred by α-(1→4) linkages (Jackson and Brennan, 2009; Mendes *et al*., 2012). Methyl groups face inwards to the center of the coil, forming a hydrophobic, cylindrical cavity, which allows accommodation of a long-chain fatty acid derivative (acyl-CoA) (Yabusaki *et al*., 1979). GG probably attaches the MGLP chains to the cytoplasmic membrane, and the reducing end is distinct from the helical conformation (Smith and Ballou, 1973; Tuffal *et al*., 1995). The structure and function of the short-chain fatty acids esterified to MGLPs is yet to be unraveled, but it has been suggested that these acyl groups stabilise the conformation of the polysaccharide chains, equivalent to the one stimulated during interaction with fatty acids (Mendes *et al*., 2012).

1.4.2 Biosynthesis of MGLPs

Two gene clusters (*Rv1208-Rv1213* and *Rv3030-Rv3037c*) were initially considered to be involved in MGLP biosynthesis in *M. tb*, until recent results indicated at least three more clusters (*Rv1326c-Rv1327c*, *Rv2417c-Rv2419c* and *Rv0126-Rv0127*) may be involved in this pathway **(Figure 1.7, Table 1.1)** (Mendes *et al*., 2012).

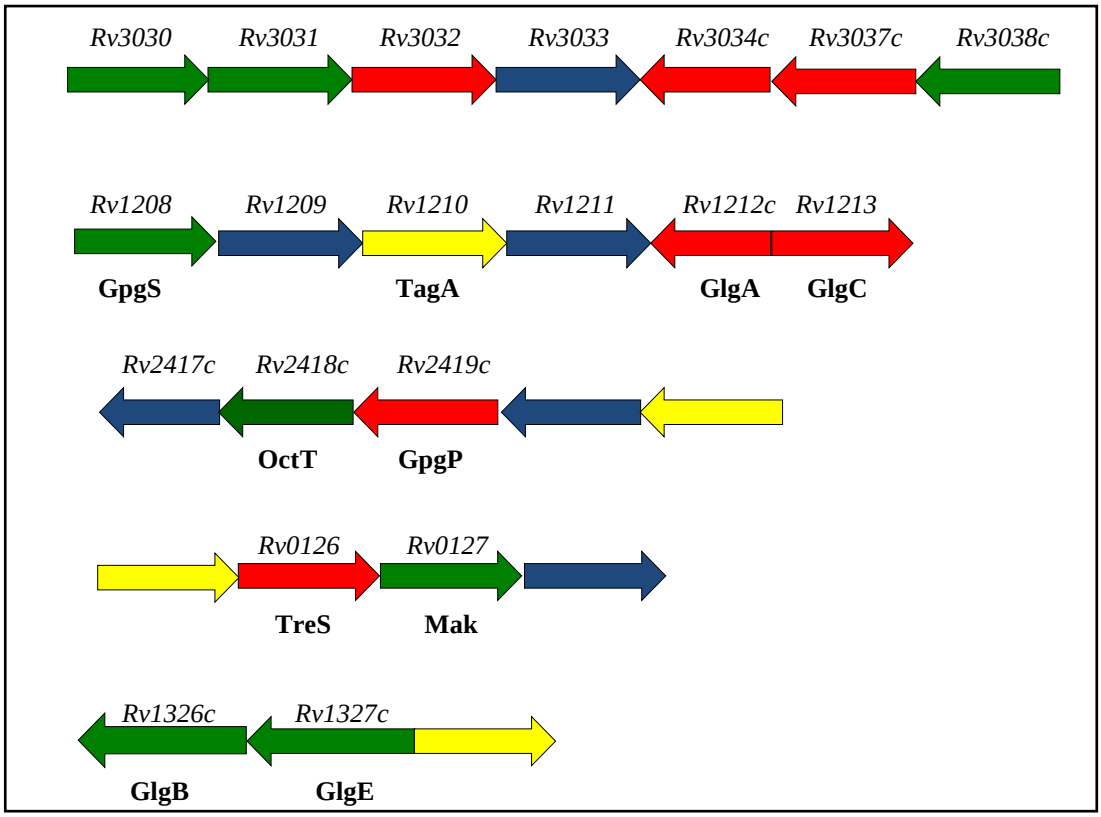


FIGURE 1.7 Gene clusters proposed to be involved in biosynthesis of MGLPs in *M. tuberculosis* H37Rv.

Shading is done as follows- green: genes essential for growth of *M. tb*; red: genes non-essential for growth of *M. tb*; blue: genes with unknown function; yellow: genes with hypothetical function (Figure adapted from Maranha *et al*., 2015; Mendes *et al*., 2012).

Gene in *M. tb* H37Rv	Protein	Proposed Function	References
Rv1208	GpgS (Glucosyl-3-phosphoglycerate synthase)	Glucosylglycerate biosynthesis	Kaur *et al*., 2009b
Rv2419c	GpgP (Glucosyl-3-phosphoglycerate phosphatase)	Glucosylglycerate biosynthesis	Mendes *et al*., 2011
	GgH* (Glucosylglycerate hydrolase)	Glucosylglycerate hydrolysis	Alarico *et al*., 2014
Rv2418c	OctT (Octanoyltransferase)	Octanoylation of GG and DGG	Maranha *et al*., 2015
Rv3031	Putative glycoside hydrolase	Probably DGG biosynthesis	Mendes *et al*., 2012; Stadthagen *et al*., 2007
Rv1326c	GlgB (Glycogen branching enzyme)	Probably DGG biosynthesis	Garg *et al*., 2007; Mendes *et al*., 2012
Rv3032	α-(1→4) glycosyltransferase	Glycosylation of the MGLP chain	Stadthagen *et al*., 2007
Rv1212c	GlgA (α-(1→4) glycosyltransferase)	Glycosylation of the MGLP chain	Stadthagen *et al*., 2007; Sambou *et al*., 2008
Rv3030	Putative SAM-dependant 6-O-methyltransferase	Methylation at 6-O position of glucoses in MGLP chain	Stadthagen *et al*., 2007
Rv3037c	Putative SAM-dependant 3-O-methyltransferase	Methylation at 3-O position of terminal glucose at non-reducing end of MGLP chain	Stadthagen *et al*., 2007
Rv3034c	Putative acetyltransferase	Esterification of glucoses of MGLP chain	Stadthagen *et al*., 2007

Table 1.1 Genes and enzymes involved in MGLP biosynthesis in *M. tuberculosis*.

* GgH enzyme is found only in fast-growing mycobacteria, and not in slow-growing species like *M. tuberculosis.*

Kamisango and colleagues proposed a model for the biosynthesis of MGLPs in which biosynthesis of the main polysaccharide chain is controlled by synchronized interplay of glycosylation, acetylation and methylation reactions, and occurs in a stepwise fashion from reducing to non-reducing end through sequential glycosylation-methylation reactions **(Figure 1.8)** (Kamisango *et al*., 1987). However, it is not yet known how the synthesis of the chain is terminated (Jackson and Brennan, 2009).

Two genes, *Rv1212c* and *Rv3032,* which encode for α-(1→4) glycosyltransferase, are associated with glycosylation and elongation of the MGLP chain in *M. tb* (Stadthagen *et al*., 2007). Knock-out mutants of *Rv3032* led to production of low levels of MGLPs, but not critically low, and its activity was shown to be partially compensated by *Rv1212c,* which is basically involved in biosynthesis of glycogen and capsular α-glucan (Sambou *et al*., 2008; Stadthagen *et al*., 2007). However, the bacilli did not recuperate after double disruption of these genes, thus implying that at least one functional gene is required for its survival (Sambou *et al*., 2008; Stadthagen *et al*., 2007). *Rv3032* was also shown to be involved in glycogen and capsular α-glucan biosynthetic pathways, though not as extensively as *Rv1212c* (Sambou *et al*., 2008). It is not yet known which genes are responsible for carrying out the β-(1→3) glycosyl-branching.

Ballou and co-workers identified and characterized a soluble protein from extracts of *M. phlei* that methylates at 6-O and 3-O positions of some glucosyl units in MGLPs and partially acetylated α-(1-4)-D-gluco-oligosaccharides, using SAM as the methyl group donor for the methyltransferase reactions (Ferguson and Ballou, 1970). Methylation at a particular position on the sugar moiety is dependent on its degree of acetylation, *i.e.* methylation occurs only at 6-O position at low levels of acetylation, whereas high levels of acetylation inhibits methylation at 6-O position, and some methylation then occurs at 3-O position (Grellert and Ballou, 1972). Methyltransferases involved in MGLP synthesis in *M. tb* are encoded by the genes *Rv3030* and *Rv3037c* (detailed in section 1.4.4) (Stadthagen *et al*., 2007).

Tung and Ballou demonstrated that the acylation of the MGLP chain is carried out by a membrane-associated acyltransferase that catalyzes the transfer of acetyl, propionyl, isobutyryl, octanoyl and succinyl groups from their respective acyl-CoA derivatives, though at different rates (Tung and Ballou, 1973). This advocates possibility of more than one acyltransferase involved in synthesis. Nevertheless, a novel acyltransferase in mycobacteria

has been identified, an octonoyltransferase- OctT (*Rv2418c*) that transfers an octanoate moiety to the C6-OH position of glucose of GG and DGG (Maranha *et al*., 2015). It has been also proposed that if the same octanoylation of GG at C6-OH position occurs *in vivo*, it can restrain the synthesis of MGLP chain, as this is the point where the second glucose of DGG attaches, and subsequently the chain is progressed (Maranha *et al*., 2015; Stadthagen *et al*., 2007).

During nitrogen deficiency, GG gets accumulated, and when the nitrogen levels are replenished, GG is hydrolyzed by GG hydrolase (GgH), an enzyme present only in the fast-growing species, to produce glucose and glycrate, and which possibly regulate the MGLP biosynthesis in rapidly-growing mycobacteria (Alarico *et al*., 2014).

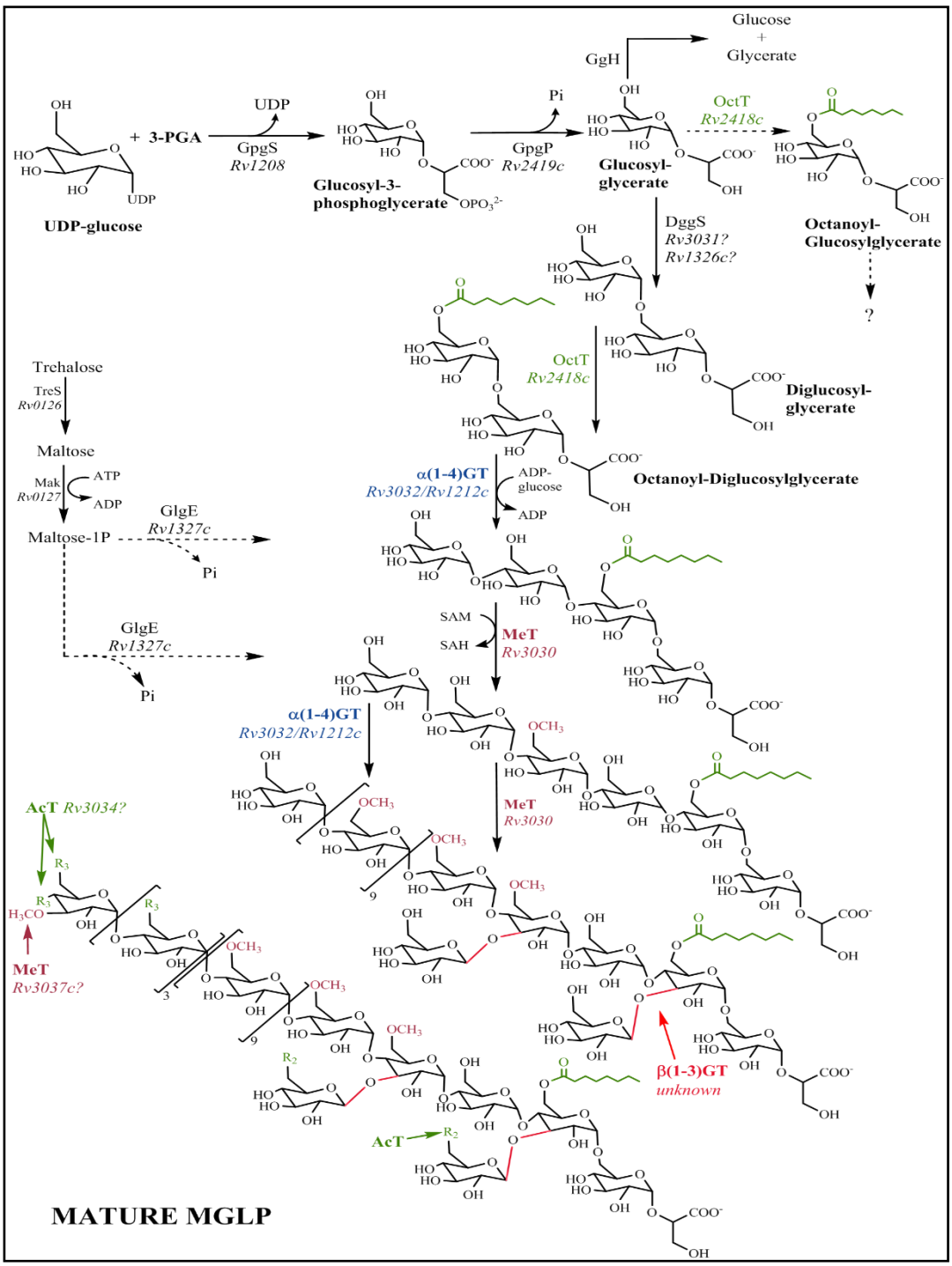


FIGURE 1.8 Proposed pathway for biosynthesis of MGLPs.

Question marks represent probable gene function. Dashed arrows represent hypothetical steps in biogenesis. α–(1→4) glycosylations (blue) are responsible for elongation of the MGLP main chain whereas β–(1→3) glycosylations (red) add branching glucoses. GT: glycosyltransferase; MeT: methyltransferase (maroon); AcT: acyltransferase (green); R2 and R3 are acyl groups (green) and methyl groups are in maroon. 3-PGA: D-3-phosphoglyceric acid; SAM: S-adenosylmethionine; SAH: S-adenosylhomocysteine; GpgS: glucosyl-3-phosphoglycerate synthase; GpgP: glucosyl-3-phosphoglycerate phosphatase; GgH: glucosylglycerate hydrolase (found only in fast-growing mycobacteria); OctT: octanoyltransferase; DggS: diglucosylglycerate synthase; TreS: trehalose synthase; Mak: maltokinase; GlgE: maltosyltransferase. (Figure adapted from Alarico *et al.*, 2014; Maranha *et al.*, 2015; Mendes *et al.*, 2012).

1.4.3 Biological Role of MGLPs

The majority of *in vitro* studies to decipher the physiological functions of MGLPs were performed in Konrad Bloch's laboratory (Flick and Bloch, 1974; Ilton *et al*., 1971; Machida and Bloch, 1973). A salient feature of PMPSs is their capability to form a stable 1:1 ratio complex with fatty-acyl and acyl-CoA of lengths ranging from C_{16} to C_{22}, and shortening or lengthening of the fatty-acyl chains notably lowers the affinity of interaction and never reaches a 1:1 molar ratio (Machida and Bloch, 1973). They help in sequestering the newly synthesized products of fatty acid synthase-I (FAS-I) and regulate the activity rate of the mycobacterial FAS-I complex by alleviating product inhibition and allowing continuous synthesis of acyl-CoAs (Bloch, 1977; Machida and Bloch, 1973). These polysaccharides significantly decrease the K_m of FAS-I for acyl- CoA substrates, thus enabling synthesis at low substrate concentrations, and a higher enzyme turnover (Flick and Bloch, 1974; Ilton *et al*., 1971; Machida and Bloch, 1973).

PMPSs prevent metabolism disruption due to fatty acid accumulation, as free fatty-acid chains (C_{16}-C_{22}) inhibit activity and inactivate several enzymes (Kawaguchi and Bloch, 1974; Yabusaki *et al*., 1979). Mycobacterial PMPSs can also serve as fatty acid carriers (Mendes *et al*., 2012; Yabusaki *et al*., 1979) and safeguard the long chain acyl-CoAs from degradation by the action of cytoplasmic lipolytic enzymes (Yabusaki and Ballou, 1979).

Mycobacteria adapt to temperature changes by rapidly altering the length, level of saturation and chain branching of mycolic acids in the outer membrane (Baba *et al*., 1989; Stratton *et al*., 2002; Toriyama *et al*., 1980). It has been hypothesized that MGLPs help mycobacteria in thermal adaptation, as a study indicated that the growth rate of *M. smegmatis* mutants, synthesizing very low levels of MGLPs, was severely affected at high temperatures (39-42°C) (Stadthagen *et al*., 2007).

1.4.4 SAM-Dependent Methyltransferases in MGLP Biosynthesis

The gene cluster *Rv3030-Rv3037c*, considered to participate in MGLPs biosynthesis, contains two putative SAM-dependent methyltransferase genes, *Rv3030* and *Rv3037c* (Stadthagen *et al*., 2007), belonging to methyltransferase family-11 and methyltransferase family-26, respectively (identified by Pfam) (Punta *et al*., 2012). The former catalyses 6-O-methylation of glucoses of MGLP, while the latter is proposed to catalyse 3-O-methylation of the terminal glucose at non-reducing end of the MGLP chain (Jackson and Brennan, 2009; Mendes *et al*., 2012). Stadthagen and colleagues created the knock-out mutants of Rv3030 orthologue in *M. smegmatis* (earlier annoted as *MSMEG2349,* now annotated as *MSMEG_2350*) and found that the disruption of this gene severely impaired the growth of the mutant at high temperatures (39-42°C) and the mutant completely obliterated the MGLP activity as compared with the wild-type (Stadthagen *et al*., 2007). This indicates that *Rv3030* orthologue in *M. smegmatis* play a pivotal role in the biosynthesis of MGLPs.

An additional SAM dependent methyltransferase gene, *Rv3038c*, found adjacent to *Rv3037c* **(Figure 1.7)**, is a member of methyltransferase family-11 (identified by Pfam) (Punta *et al*., 2012) and contains LysR family signature but shows no similarity to known LysR family members (as given in TubercuList database) (Lew *et al*., 2011).

Rv3030 and *Rv3038c* are regarded essential, whereas *Rv3037c* is non-essential for *M. tuberculosis* growth by high-density mutagenesis (Griffin *et al*., 2011; Sassetti *et al*., 2003), however, the latter is required for survival of the pathogen in primary murine macrophages, as demonstrated by transposon site hybridization (TraSH) in *M. tuberculosis* H37Rv (Rengarajan *et al*., 2005). These genes are essential, either for growth or survival of *M. tuberculosis*, and are promising drug targets (Griffin *et al*., 2011; Mendes *et al*., 2012; Rengarajan *et al*., 2005; Sassetti *et al*., 2003; Stadthagen *et al*., 2007).

1.5 A Possible Link between Trehalose Metabolism, MGLP Chain Elongation and Biosynthesis of Capsular α-glucan and Glycogen

Maltose is produced from trehalose in a reaction catalyzed by TreS (*Rv0126*) (described afore in section 1.3.1), which is then phosphorylated by maltokinase (Mak or Pep2, *Rv0127*) to produce maltose-1-phosphate (M1P), the substrate of maltosyltransferase (GlgE, *Rv1327c*) that elongates glycogen chains, and possibly other α-(1→4) glucose polymers such as the capsular a-glucan or the intracellular MGLPs (Elbein *et al*., 2010; Kalscheuer *et al*., 2010; Miah *et al*., 2013). Gene *Rv1326c*, encoding for α-(1→6) glycogen branching enzyme (GlgB), is present upstream of *Rv1327c*, and both genes are present in the gene cluster involved in glycogen synthesis (Elbein *et al*., 2010; Kalscheuer *et al*., 2010). It has been hypothesized that GlgB has an involvement in the synthesis of DGG, an early intermediate for MGLP chain synthesis (Elbein *et al*., 2010; Jackson and Brennan, 2009) (section 1.4.2). It has been proposed that under normal growth conditions, glycogen would be synthesized by the GlgA-GlgB *(Rv1212c-Rv1326c)* pathway, while during heat-stress, excess trehalose would be used for glycogen production via TreS-Mak-GlgE (Elbein *et al*., 2010). The putative connection of the gene clusters *Rv0126-Rv0127* and *Rv1326c-Rv1327c* in the MGLP biogenesis, offers a possibility of utilizing trehalose via Tres-Mak-GlgE and glycogen biosynthetic genes as alternative routes for biosynthesis and elongation of MGLP chains, though it still needs to be experimentally validated **(Figure 1.8)** (Elbein *et al*., 2010; Jackson and Brennan, 2009; Kalscheuer *et al*., 2010).

1.6 Drug Development Against Tuberculosis: Combating the Old Enemy

The research for development of drugs against TB has seen groundbreaking advancements, but has faced severe issues and challenges. Despite the success of highly efficacious oral four-drug (isoniazid, ethambutol, rifampicin and pyrazinamide) anti-tubercular combinatorial treatment, introduced almost 60 years ago, TB continues to strikingly increase the morbidity and mortality rates (Ma *et al*., 2010; Stehr *et al*., 2014). After the discovery of rifampicin in the 1960s, a long period passed without witnessing a new formulation to treat this old enemy,

up until last decade wherein several novel treatment options have emerged, now being studied in pre-clinical and clinical trials (Mdluli *et al*., 2015; Sloan *et al*., 2013; Zumla *et al*., 2014; Zuniga *et al*., 2015). Moreover, drug-resistance of the bacilli and HIV-TB co-pandemic has substantially pressurised the healthcare sector, necessiating the need for new and more potent treatment regimens.

The typical first-line treatment against drug-susceptible TB involves combination of rifampicin, isoniazid, ethambutol and pyrazinamide given for a minimum period of 6-9 months, with an 'intensive phase' of 4 drugs (rifampicin, isoniazid, pyrazinamide and streptomycin or ethambutol) for the first two months, followed by a four month 'continuation phase' of two drugs (rifampicin and isoniazid) **(Figure 1.9, Table 1.2)** (Piccaro *et al*., 2013; Zumla *et al*., 2014). However, the current treatment is long, complex and has severe side-effects, which leads to non-adherence, and as a consequence, there is a high risk of failure and relapse, continuous spread of infection and resistance to the antibiotics (Zumla *et al*., 2014). Management of drug resistant TB, multi-drug (MDR-TB) and extreme-drug (XDR-TB), necessitates further administration of second-line (aminoglycosides, polypeptides, fluoroquinolones) and third-line (macrolides, linezolid, thioacetazone) drugs, which prolongs the treatment and makes it even more arduous to manage **(Table 1.2)** (Chan and Iseman, 2008; Gandhi *et al*., 2010; Koul *et al*., 2011; Riccardi *et al*., 2009; Zumla *et al*., 2012). The problem with present therapy is further augmented by incompatibility of anti-TB drugs, particularly rifampicin, with many anti-retroviral drugs, thus, making it inconvenient to use for HIV-TB co-infection (Aaron *et al*., 2004; Breen *et al*., 2006; Centers for Disease Control and Prevention (CDC), 2002; McIlleron *et al*., 2007; Nannelli *et al*., 2008).

Overall, there is a pressing need for development of novel therapeutics that are more potent to reduce treatment duration, inhibit new targets to be effective against MDR- and XDR-TB, are compatible with anti-retroviral treatment, have no antagonistic drug-drug interactions, and have fully assessed safety profile (Wallis *et al*., 2016; Zumla *et al*., 2014; Zuniga *et al*., 2015).

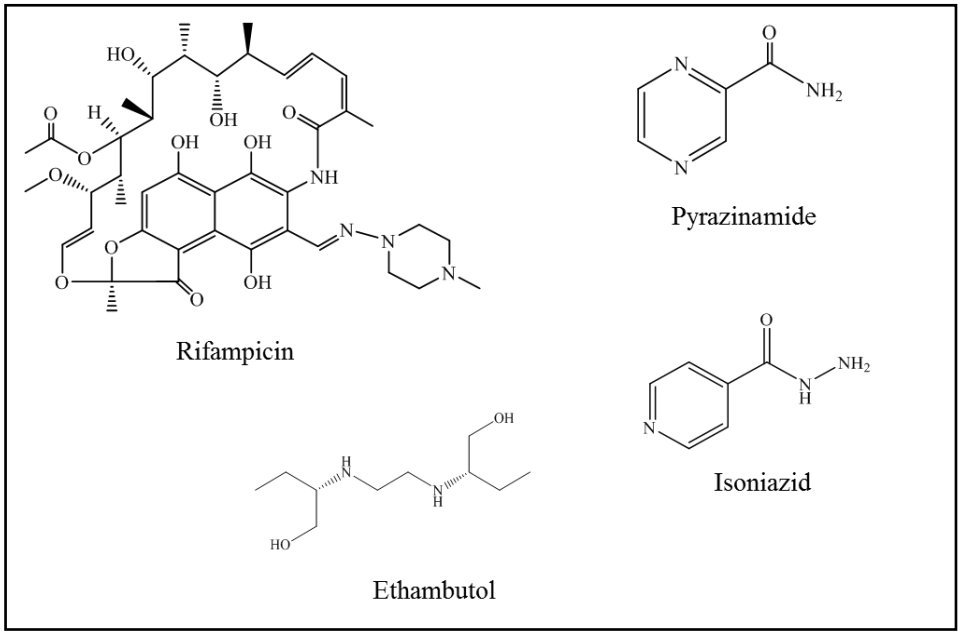


FIGURE 1.9 Chemical structures of current first-line drugs used for treating drug-susceptible TB.

Drug (Chemical class)	Year of Discovery	Target	Mode of Action	References
First-Line Drugs				
Rifampicin (Rifamycin)	1963	rpoB (β-subunit of RNA polymerase)	Inhibition of transcription	(Dickinson and Mitchison,1981; Węgrzyn *et al.*, 1998)
Isoniazid (Hydrazide)	1952	inhA (Enoyl-[acyl-carrier-protein] reductase)	Inhibition of mycolic acid biosynthesis	(Bollela *et al.*, 2016; Rozwarski *et al.*, 1999)
Ethambutol (Ethylenediamine)	1961	embA, embB, embC (indolylacetylinositol arabinosyltransferase)	Inhibition of arabinogalactan and LAM biosynthesis	Goude *et al.*, 2009
Pyrazinamide (Amides)	1954	pncA (Pyrazinamidase, 30s ribosomal subunit)	Inhibits translation, trans-translation, and affects the pH balance of the cytoplasm	(Scorpio and Zhang, 1996; Scorpio *et al.*, 1997; Sun *et al.*, 1997)

(cont.)

Drug (Chemical class)	**Year of Discovery**	**Target**	**Mode of Action**	**References**
Second-Line Drugs				
Streptomycin (Aminoglycoside)	1944	Ribosomal S12 gene and 16S rRNA components of 30S ribosomal subunit	Inhibition of protein synthesis	(Finken *et al.*, 1993; Ruusala and Kurland, 1984)
Kanamycin (Aminoglycoside)	1957	rrs (30S ribosomal subunit)	Inhibition of protein synthesis	(Alangaden *et al.*, 1998; Sirgel *et al.*, 2012)
Amikacin (Aminoglycoside)	1972	rrs (30S ribosomal subunit)	Inhibition of protein synthesis	(Alangaden *et al.*, 1998; Sirgel *et al.*, 2012)
Ofloxacin (Fluoroquinolone)	1980	gyrA, gyrB (DNA gyrase and DNA topoisomerase)	Inhibition of DNA supercoiling	(Chien *et al.*, 2016)
Para-amino salicylic acid (Salicylic acid derivatives)	1948	folP (Dihydropteroate synthase)	Inhibition of thymidylate synthase, folate biosynthesis and interference with iron uptake	(Donald and Diacon, 2015)
Ethionamide (Thioamide)	1961	inhA (Enoyl-[acyl-carrier-protein] reductase)	Inhibition of mycolic acid biosynthesis	(DeBarber *et al.*, 2000; Wang *et al.*, 2007)
Cylcoserine (Amino acid analogues)	1955	ddlA (D-alanine racemase and ligase)	Inhibition of peptidoglycan biosynthesis	(Bruning *et al.*, 2011)

Table 1.2 Current drugs for treatment of tuberculosis.

LAM: lipoarabinomannan

After a long stagnant phase, the drug-development pipeline for TB has been filled with several novel, diverse and promising candidates in the last decade. The chemical entities, either the repurposed drugs or new compounds, are at different stages of development, as depicted in **Figure 1.10,** with description of drugs in **Figure 1.11** and **Table 1.3**. There is also a huge need for drugs that can kill *M. tb* in latent phase and treat latent TB infection (Santivañez-Veliz *et al*., 2016; Stuurman *et al*., 2016).

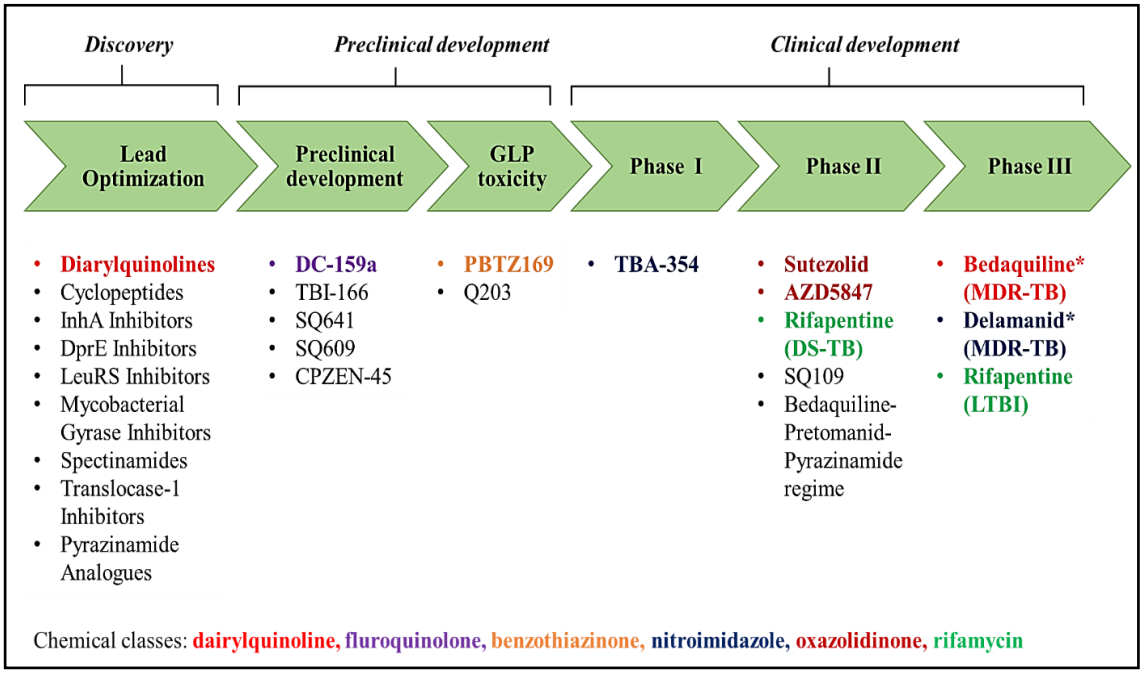
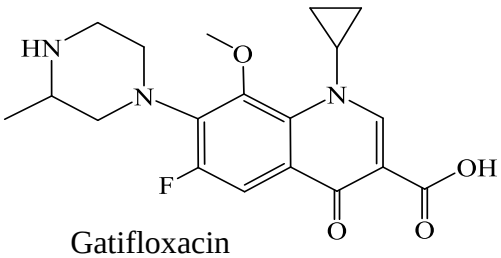


FIGURE 1.10 TB drug development pipeline.

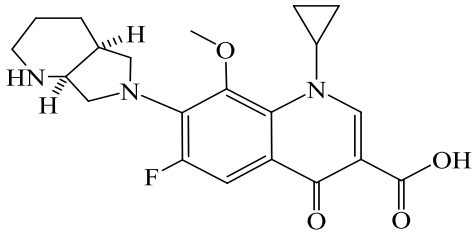
Global pipeline of chemical entities being discovered or developed for the treatment of TB (as in August 2015). GLP: good laboratory practice; DS-TB: drug-susceptible tuberculosis; MDR-TB: multidrug-resistant tuberculosis; LTBI: latent tuberculosis infection; InhA: enoyl CoA reductase; LeuRS: leucyl-tRNA synthetase. Details of ongoing projects can be found on the Working Group on New TB Drugs-Pipeline website.
* along with Optimized Background Regimen
(Figure adapted from Global tuberculosis report, WHO, 2015).

A. Repurposed Drugs

Fluoroquinolones

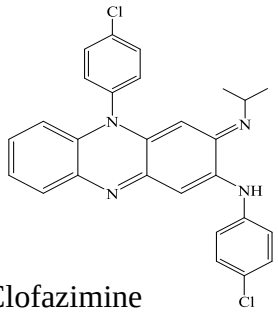


Gatifloxacin



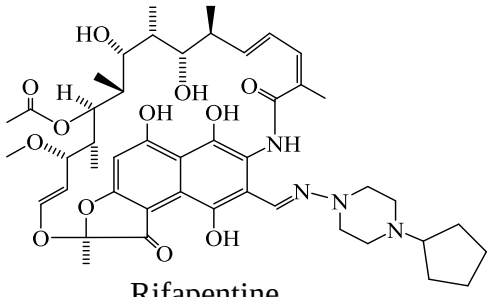
Moxifloxacin

Riminophenazines



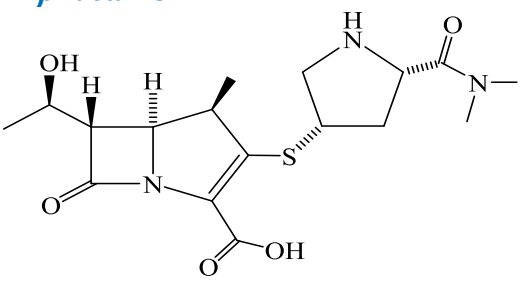
Clofazimine

Rifamycins

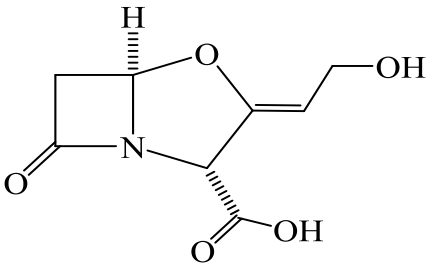


Rifapentine

β-lactams

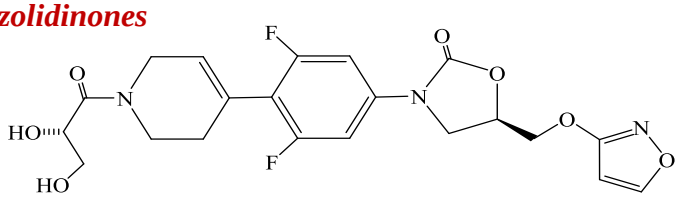


Meropenem

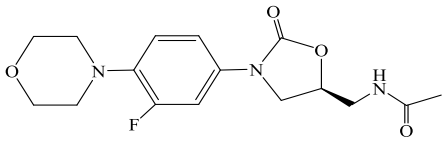


Clavulanate

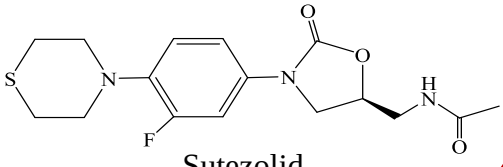
Oxazolidinones



AZD-5847



Linezolid



Sutezolid

(cont.)

B. New Chemical Entities

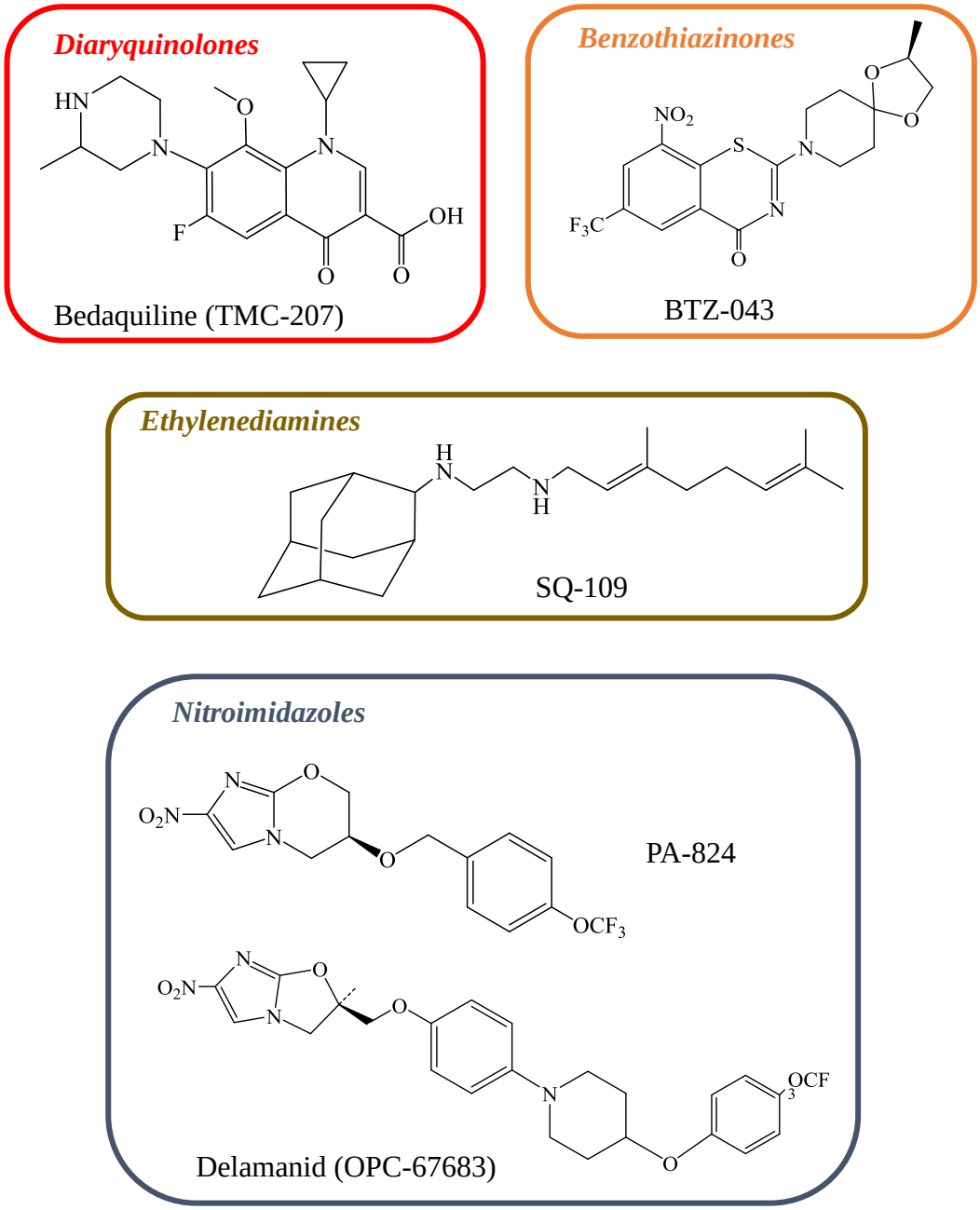


FIGURE 1.11 Chemical structures of repurposed and novel TB drugs.

A. Drugs that are used for treatment of other diseases, now being assessed for treating tuberculosis.
(Figure adapted from Zumla *et al.*, 2013).

B. Novel drugs developed specifically for treating tuberculosis.
(Figure adapted from Stehr *et al.*, 2014).

Chemical Class	Drugs	Properties	References
Fluoroquinolones	• Gatifloxacin • Moxifloxacin	• Target DNA gyrase, and inhibits DNA synthesis	(Nuermberger *et al*., 2004; Zumla *et al*., 2013)
Riminophenazines	• Clofazimine	• Leprosy drug repurposed for treating TB • Targets the outer membrane	(Dey *et al*., 2013; Verma *et al*., 2013)
Rifamycins	• Rifapentine	• Targets β-subunit of RNA polymerase, and inhibits transcription • Longer half-line than rifampicin	(Benator *et al*., 2002; Rosenthal *et al*., 2007; Steingart *et al*., 2011; Sterling *et al*., 2011)
β-lactam antibiotics	• Combination of Meropenem and Clavulanate	• Clavulanate inhibits activity of β-lactamase (BlaC) • Meropenem inhibits activity of DD-carboxy-peptidase • Alter PG biosynthesis	(De Lorenzo *et al*., 2013; Hugonnet *et al*., 2009; Kumar *et al*., 2012)
Oxazolidinones	• Linezolid • Sutezolid (PNU-100480) • AZD-5847	• Target the 23S rRNA in the 50S ribosomal subunit, and inhibit protein synthesis	(Ashtekar *et al*., 1991; Fortún *et al*., 2005; Sotgiu *et al*., 2012; Wallis *et al*., 2010, 2011, 2014)
Diarylquinolines	• Bedaquiline	• Targets the c-subunit of ATP synthase, and inhibits energy production • Equally active against dormant and active bacilli	(Andries *et al*., 2005; Cox and Laessig, 2014; Diacon *et al*., 2009; Huitric *et al*., 2010; Koul *et al*., 2007, 2008; Pym *et al*., 2016)

(cont.)

Chemical Class	Drugs	Properties	References
Benzothiazinones	• BTZ-043	• Targets the flavoenzyme DprE1, and inhibits AG and LAM biosynthesis • Nanomolar bactericidal activity *in vitro* and *ex vivo* • No antagonism with current TB drugs or the candidates	(Lechartier *et al*., 2012; Makarov *et al*., 2009; Neres *et al*., 2012; Trefzer *et al*., 2012)
Nitroimidazoles	• PA-824 • Delamanid (OPC-67683)	• Prodrugs, activated by F_{420}-deazaflavin-dependent nitroreductase (Ddn) • PA-824 inhibits mycolic acid synthesis and poisons respiratory chain • Delamanid inhibits mycolic acid synthesis and kills intracellular bacilli	(Diacon *et al*., 2011; Gler *et al*., 2012; Manjunatha *et al*., 2009, 2006; Matsumoto *et al*., 2006; Singh *et al*., 2008; Skripconoka *et al*., 2013)
Ethylenediamines	• SQ-109	• Targets MmpL3, and inhibits mycolic acid biosynthesis	(Heinrich *et al*., 2015; Lee *et al*., 2003; Tahlan *et al*., 2012)

Table 1.3 Repurposed and novel drugs for treatment of tuberculosis.

DrpE1: decaprenylphosphoryl-beta-D-ribose 2′-epimerase
AG : arabinogalactan
LAM : lipoarabinomannan
PG : peptidoglycan

The biology of *M. tb* that is advantageous for its persistant survival and development of drug-resistance, coupled with the complexiety of the human disease, imposes several hinderances in the drug discovery pathway. Although the anti-TB drug development pipeline has witnessed significant progress in the last decade, it is still limited. The treatment of TB is based on combinatorial administration of drugs to prevent drug-resistance, and thus, the drugs in the pipeline are still being trialled with the existing drugs to shorten or improve the therapy time (Zumla *et al*., 2014). The animal models used for evaluating the lead compounds before the clinical trials are not accurate representatives of the pathogenesis of the human disease, thereby making it difficult to predict the efficacy of the drug (Young, 2009). Although several agents are present in the pipeline, it would take likely another decade to fully develop and execute for clinical development. There is a compelling need for new chemical moieties as only few drugs in the pipeline are currently in the clinical development phase.

An alternate approach for the TB therapy is the adjunct immunotherapy or host-directed therapy, which directs to strengthen the host immune system, and is being considered as a supplement to the drug therapy to shorten the treatment time, prevent re-occurrence of the disease, and improve success rate of treating MDR-TB (Uhlin *et al.*, 2012; Wallis *et al*., 2016; Zumla *et al*., 2015a). This is necessary to eliminate the active bacilli or contain the infection in latent form, protect the lungs, prevent and repair damage to the organ tissue, and avert resistance (Wallis and Hafner, 2015; Zumla *et al*., 2015b). The adjunct therapy includes: immunotherapy with interleukin-2, interferon-ɣ and interleukin-7; cell based therapy with mesenchymal stromal cells; drugs used for treatment of other diseases and do not target *M. tb* directly (for instance, metformin that is used for treating diabetes promotes macrophage autophagy, anti-cancer drug imatinib affects macrophage function); micronutrients like vitamin-D that have immuno-modulatory effects; and therapeutic vaccines (Alangari *et al*., 2011; Rosenzweig *et al*., 2006; Silva *et al*., 2016; Uhlin *et al*., 2012; Wallis and Hafner, 2015; Wallis *et al*., 2016; Zumla *et al*., 2015b, 2015a).

1.7 Objectives and Approaches of the Doctoral Study

Whilst much improvement has been made in the development of novel therapeutics to treat TB, there is a continual need to identify new potential drug targets, and investigate new therapeutic agents and treatment regimens, for more effective treatment and management of the infection. It is clear from the review of literature above that mycobacterial pathogenesis is closely related to its cellular envelope, and that MGLPs and trehalose play a critical role in defining the architecture of the cell wall, bacterial growth and survival. Therefore, it is imperative to explore these pathways, which offer a multitude of novel and essential enzymes that can be targeted to combat the *M. tuberculosis* infection.

Design of therapeutics for treatment of a disease relies on understanding the basic molecular mechanisms underpinning it. The primary objective of this doctoral study is to gain knowledge about the target proteins by their structural and functional characterisation. The information gained creates a foundation and may be used to carry out an early-stage drug discovery, to identify chemical fragments that bind to proteins and modulate its function, and that could be developed into new lead inhibitors for novel anti-TB treatment **(Figure 1.12)**.

The targets investigated in this study are: Rv3030, Rv3037c and Rv3038c-SAM dependent methyltransferases involved in MGLP biogenesis; and OtsA involved in trehalose biosynthesis. The rationale behind selection of these targets was to unearth new targets that are currently not being investigated for structural and drug discovery studies, and thus, offer a novel dimension in targeting the disease mechanistic. Essentiallity of these enzymes, either for growth or for survival of the bacilli, and lack of human orthologues, make them attractive targets to study. A pre-requisite for structure determination and drug-discovery studies is production of tractable amounts of protein, purified to homogeneity. Detailed experimental analyses of these targets revealed feasibility of only two targets, Rv3030 and OtsA, which will be described in detail in the following chapters. Initial experimental data on the remaining targets, Rv3037c and Rv3038c, is presented in the annex.

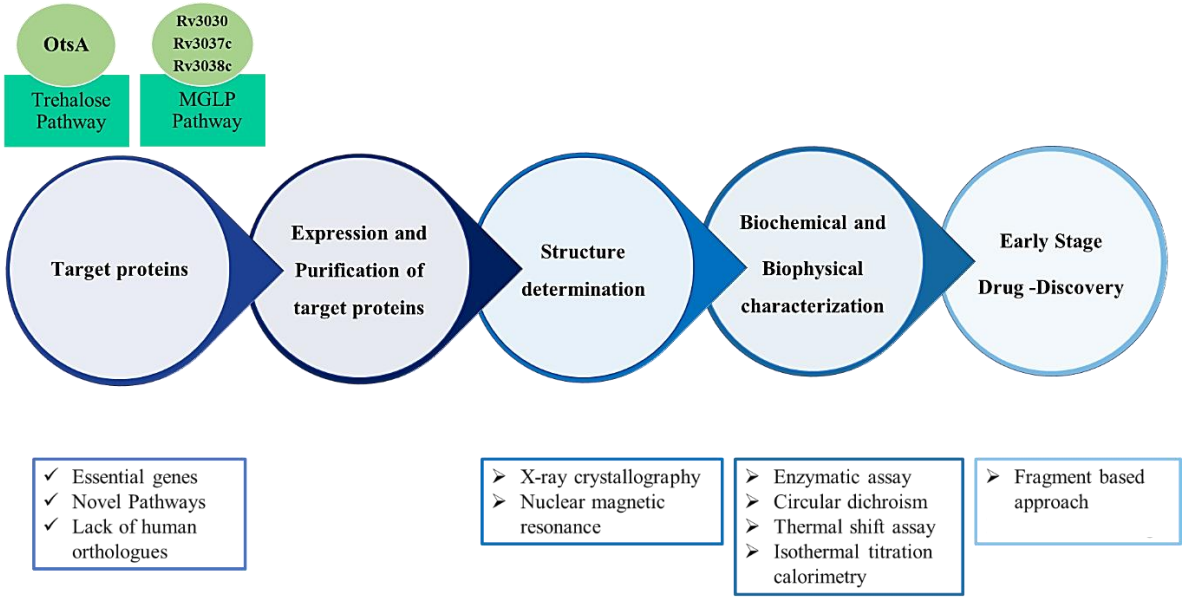


FIGURE 1.12 Objectives and approaches of the doctoral study.

Organisation of Thesis

This thesis describes in detail, the work done on the targets OtsA and Rv3030.

Each corresponding chapter on these targets consists of a summary, an introduction, the methodology, results and discussion.

- **Chapters 2** details the work on **Target-1: OtsA**, trehalose 6-phosphate synthase involved in trehalose biosynthesis.
- **Chapters 3 and 4**, documents work on **Target-2: Rv3030**, methyltransferase involved in MGLP biogenesis.

Chapter 5 presents final concluding remarks and future perspectives.

Additional supporting information is present in **Appendices I- IV**.

Preliminary experimental data on targets **Rv3037c and Rv3038c** is presented in **Appendix-V**.

CHAPTER-2

Functional and Structural Studies, and Probing the Druggability of *M. thermoresistibile* OtsA

2.1 Summary

This chapter focuses on studies of mycobacterial trehalose-6-phosphate synthase, through the functional and structural characterisation of the enzyme, which form the basis for a possible structure-guided drug discovery campaign for antibacterials targeting the *Mycobacterium tuberculosis* enzyme. As the *Mycobacterium tuberculosis* enzyme could not be expressed in the soluble fraction, other mycobacterial orthologues were investigated. The orthologue from *Mycobacterium thermoresistibile* was expressed in *E. coli*, and the recombinant enzyme was successfully purified and crystallised. The best crystals diffracted to 2.18 Å, and crystallised in cubic symmetry. The three-dimensional structures in both apo and ligand-bound states were determined. The previously reported structures of OtsA from *E. coli* are of protein-ligand complexes (Gibson *et al*., 2002a, 2004). This work describes the first mycobacterial OtsA structure. The apo-structure of *M. thermoresistibile* OtsA was also the first structure of trehalose-6-phophate synthase in the apo-form, although subsequently the apo-enzyme structure from *B. xenovorans* also has been determined (February, 2016). Biochemical characterisation showed that the enzyme was able to utilise several glucose-donor substrates, like other mycobacterial OtsAs, but displayed highest preference for ADP-glucose, and was strictly dependent on glucose-6-phosphate as the acceptor substrate. Moreover, the activity of the enzyme was shown to be regulated by feedback inhibition. The donor-substrate bound structures of protein also offer a plausible explaination for substrate preference.

This chapter also focuses on early-stage drug discovery against the mycobacterial OtsA, using *M. thermoresistibile* OtsA as the representative system. Fragment-based approaches were used for identification and validation of hits. Preliminary screening of a small-scale fragment library by thermal shift assay led to identification of several fragments that bind to the enzyme. The top hits among these, were further validated by inhibition assay. However, it was not possible to obtain fragment-bound structures with any of the fragments that were tested.

2.2 Introduction

Trehalose-6-phosphate synthase (OtsA) is a glycosyltransferase that catalyses the transfer of glucose from a nucleotide-glucose donor to an α-anomer of glucose-6-phosphate, the acceptor substrate, to form trehalose-6-phosphate. Glycosyltransferases are enzymes that catalyse formation of glycosidic bonds using either a nucleotide- or a lipid-dependent sugar donor (Varki *et al*., 2009). During the anomeric reaction, the net stereochemistry of the sugar-donor substrate can be retained or inverted, and as a result, the mode of catalysis is different (Lairson *et al*., 2008). The inverting glycosyltransferases employ a straightforward single substitution S_N2 mechanism with an oxocarbenium-ion transition state (Lairson *et al*., 2008). On the other hand, the catalytic mechanism by retaining glycosyltransferases is ambiguous, and probably involves a Koshland double-displacement mechanism, as in glycoside hydrolases, or an S_Ni 'internal return' process (Lairson *et al*., 2008). The reaction catalysed by OtsA is proposed to proceed through same-face S_Ni process, with the net retention of the stereochemistry of the sugar-donor substrate. In this reaction, the nucleotide-phosphate group of the donor leaves and the nucleophile, glucose-6-phosphate, attacks from the same face, while the other face of the glucose-donor substrate is guarded by the enzyme. However, it remains unclear whether the reaction occurs in a concerted manner, or in a stepwise fashion involving formation of a metastable intermediate (Errey *et al*., 2010; Lee *et al*., 2011).

Two structural classes have been identified for nucleotide-dependent enzymes, GT-A and GT-B, both of which have the classic Rossmann fold. The enzymes with GT-A fold have two adjoining Rossmann domains, whereas the GT-B topology has two flexibly-linked Rossmann domains facing each other (Bourne and Henrissat, 2001; Lairson *et al*., 2008). There is no association between the catalytic mechanism and the overall fold, as retaining and inverting glycosyltransferases belong to both structural classes (Coutinho *et al*., 2003). The structures of *E. coli* OtsA revealed that it is the first retaining glycosyltransferase whose structure possesses the GT-B topology (Gibson *et al*., 2002a, 2004). Distinct folds are now evident for the soluble domains of the lipid-dependent glycosyltransferases. For example, the lipid-dependent glycosyltransferase (GT51) synthesising peptidoglycan has a bacteriophage lysozyme-like fold, instead of the Rossmann fold (Lovering *et al*., 2007; Yuan *et al*., 2007).

The trehalose-6-phosphate synthase reaction was first demonstrated in the cell-free extracts of *S. cerevisiae* (Cabib and Leloir, 1958), and has been observed in an assorted set of

organisms: mycobacteria (Pan *et al*., 2002, 2004), *E. coli* (Gibson *et al*., 2002b), *D. discoideum* (Roth and Sussman, 1966), silk moths (Murphy and Wyatt, 1965), locusts (Candy and Kilby, 1959) and *Arabidopsis thaliana* (Vogel *et al*., 2001). Mycobacteria produce trehalose through multiple pathways, and trehalose-conjugates play a vital role in defining the architecture of the cell wall and the virulence of the bacilli. The prime OtsAB pathway for trehalose biosynthesis utilizes trehalose-6-phosphate synthase (Tps/OtsA) and trehalose-6-phosphate phosphatase (Tpp/OtsB2) enzymes. In *M. tb*, the *otsA* (*Rv3490*) gene encodes for trehalose-6-phosphate synthase, which catalyses the first step of OtsA-OtsB2 pathway, producing trehalose-6-phosphate from NDP-glucose and glucose-6-phosphate.

Targeting trehalose biosynthetic pathway is useful for development of novel anti-TB therapies. This is evident from the fact that trehalose is important in physiology and virulence of *M. tb*, and is absent in humans. Disruption of trehalose-6-phosphate synthase leads to growth impairment or lethality in several organisms, and thus, it is a potent drug target. For example, in *D. melanogaster*, TPS is an insecticidal target because insertion in tps1 gene was lethal during early larval stage (Kern *et al*., 2012); mutants of TPS in *C. neoformans* and *C. gattii* were attenuated in virulence and thus, it is a fungicidal target (Petzold *et al*., 2006). In *M. tb*, OtsA mutants show severe impaired growth and a substantial decrease in virulence in mouse models (Murphy *et al*., 2005). This makes OtsA an attractive target to develop chemical moieties that can aid in fighting against TB.

2.3 Basic Principles of Protein Crystallisation and Crystallography, and Other Techniques Deployed

2.3.1 Protein Crystallisation and Data Collection

Protein crystallisation is based on the principle that protein solution of high concentration (in mg ml^{-1}) is mixed with precipitant, which reduces protein solubility. When solution becomes supersaturated and thermodynamically metastable, spontaneous nucleation occurs and crystals may start to form **(Figure 2.1)**. Several crystallisation trials are done with different reagents, varying reagent concentrations, pH and temperature, to determine conditions that

will yield diffracting crystals. There are several crystallisation techniques, but the most common method is vapour diffusion.

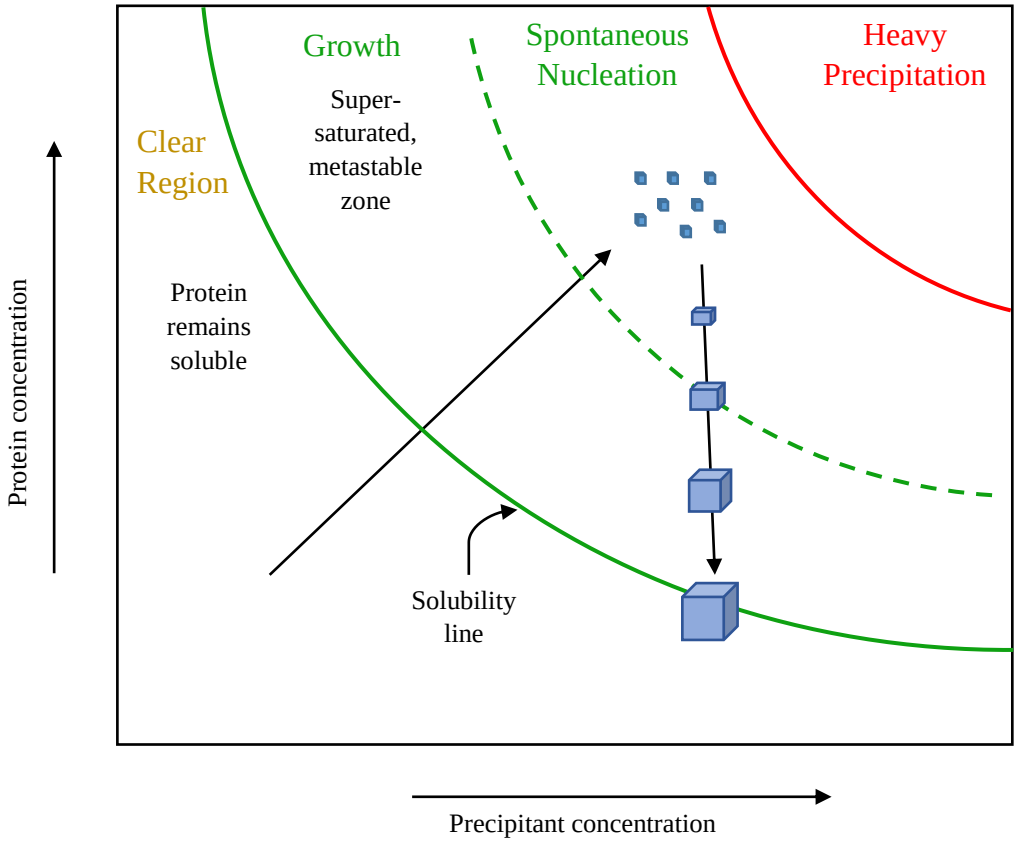


FIGURE 2.1 Phase diagram of protein crystallisation.

Drops remain clear when protein and precipitant solution are not mixed in appropriate concentrations and are unable to cross the solubility line, or when spontaneous nucleation does not occur even though drops are set with protein and precipitant in sufficiently high concentrations. If the concentrations are too high, the drops immediately precipitate. When drops reach the spontaneous nucleation region, crystal nuclei are formed and crystals grow further in the metastable growth region, till they are in equilibrium with the protein solution. This is when the crystals reach their final size and the growth terminates. (Figure adapted from Rupp, 2009).

In the vapour diffusion method, protein and crystallisation cocktail are mixed in the desired ratio, and then, due to the concentration gradient, water evaporates from the crystallisation drop into the reservoir of precipitant. This makes the protein-precipitant solution more concentrated, which may lead to supersaturation, nucleation and crystallisation. Vapour diffusion is implemented either as hanging- or sitting-drop technique **(Figure 2.2)**. In the

hanging-drop method, a protein drop is mixed with crystallisation cocktail in the reservoir and placed on siliconized cover slide. The rim of the reservoir is greased; the flipped-over cover slide is placed on the well and sealed. On the other hand, in the sitting-drop vapour drop technique, the drop sits on a shelf or post. Sitting-drop method can be used with robotic setup, whereas hanging- drop method is used for manual setup for small-scale optimisations.

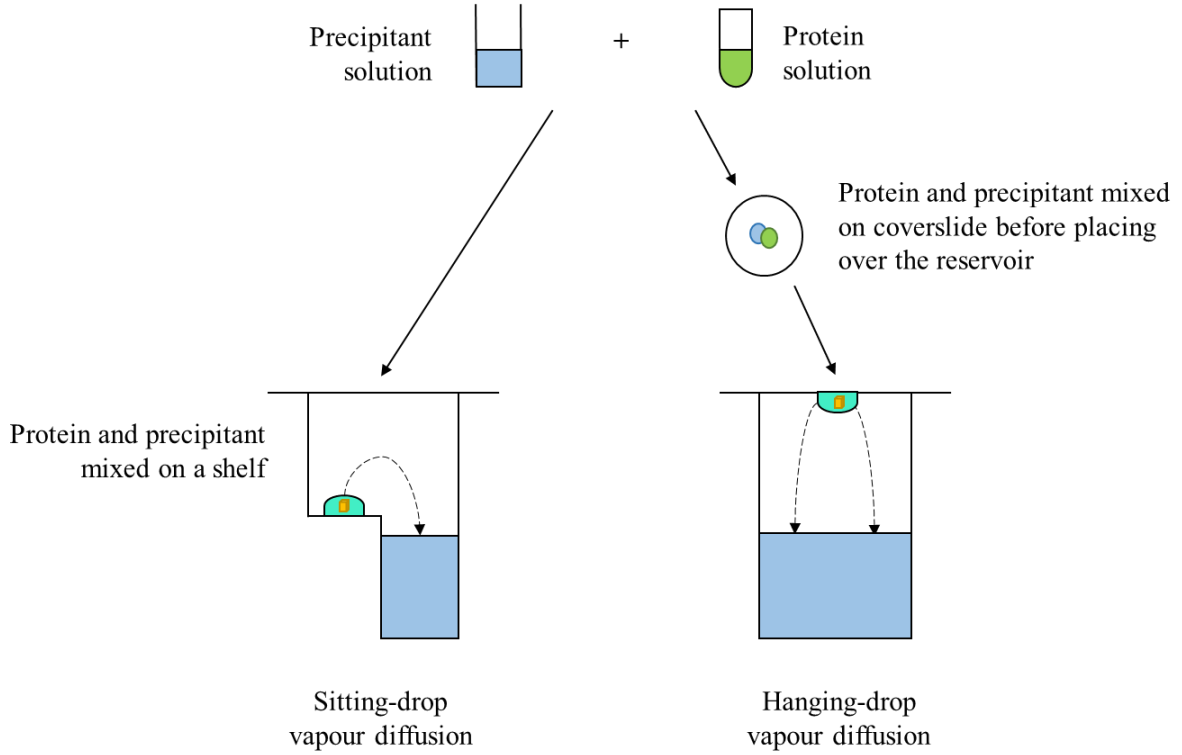


FIGURE 2.2 Diagrammatic representation of vapour-diffusion crystallisation techniques.
(Figure adapted from Rupp, 2009).

It is essential to obtain complex structure of target protein with cofactors, substrates or therapeutic lead compounds to understand the dynamics of interactions at atomic level, which take place during catalytic mechanism and protein-drug binding. The two chief routes to obtain complexes are soaking and co-crystallisation. Soaking is the technique in which the ligand is incorporated in the already grown crystal by adding the substance in the mother liquor. Co-crystallisation, on the other hand, involves incubation of protein solution with the ligand, and then crystallising the mixture together.

When crystals are grown to a suitable size, they are harvested from the drop along with the mother liquor. The harvested crystal is cryoprotected, before flash freezing in liquid N_2 and mounting it on a diffractometer for data collection. Cryoprotection plays a critical role in

protecting the crystals from radiation damage, and also prevents crystallisation of mother liquor during rapid cooling, which otherwise can be detrimental to the diffraction quality of the crystal. Routinely used cryoprotectants are glycerol, MPD, PEGs, ethylene glycol and sucrose. Screening of cryoprotectants is generally required to ensure their suitability for a specific protein. The crystals are mounted on the goniometer for X-ray data collection.

Data collection is the final experimental step before *in silico* structure determination and analysis. Most commonly, the crystal exposed to the X-ray beam is rotated about a single axis in small successive increments, and the fraction image is recorded at each rotation increment. The intensity of spots of individual images are integrated and actual reflection intensities are determined. The quality of crystal is determined by the initial diffraction image, which subsequently dictates the remaining strategic data collection process.

Several texts and reviews can be referred for the detailed understanding of protein crystallisation and developments in the field (Benvenuti and Mangani, 2007; Bergfors, 2009; Deller *et al*., 2016; Giegé, 2013; Luft *et al*., 2007; Rhodes, 2006; Rupp, 2009; Sugahara, 2014).

2.3.2 Protein Crystallography

This section summarises the basic theory of protein crystallography and techniques used in this study. Texts providing further detail and the latest developments include the classic text Protein Crystallography (Blundell and Johnson, 1976), and more recent monographs (Garman, 2014; Nannenga and Gonen, 2014; Ogata *et al*., 2015; Rhodes, 2006; Rupp, 2009).

2.3.2.1 Geometric Assembly of Protein Crystals and Space Groups

Protein crystals, which are periodic assemblies of large and flexible macromolecules, are necessary for producing X-ray diffraction. A unit cell, the basic repeating building block of a crystal, is a combination of unit lattice and molecular motifs. A unit cell is a mathematical construct defined by the lengths of three edges: a, b and c, and the angles between them: α, β

and ɣ. The asymmetric unit is the smallest part of the unit cell containing the motif that can generate the complete unit cell when space group symmetry operations are applied. A crystal lattice is the periodic construct, generated by finite repeats of unit cells. It is an imaginary three-dimensional grid, consisting of an array of points so arranged that their environments are identical.

There are different types of unit cells and are generally divided into two categories. The first category is the primitive unit cell, in which the lattice points are present only at the vertices of the unit cell and they do not have any additional internal symmetry. There are six types of primitive lattices. The second category is the centered unit cell, which contains additional lattice points to those at vertices of the cell. There are three types of centerings: face, body or base, which lead to eight centered lattices. These are collectively known as 14 Bravais lattices, and belong to one of the seven crystal systems: triclinic, monoclinic, orthorhombic, rhombohedral, hexagonal, tetragonal and cubic. A symmetry operation is the operation applied to the asymmetric unit to generate its identical copies. A space group provides complete description about the symmetry of the unit cell and its contents. It reveals information about the lattice type and the symmetry operations used. In total, there are 230 space groups, of which only 65 are possible for chiral proteins.

2.3.2.2 Reciprocal Lattice and Diffraction Basics

X-rays are electromagnetic radiations of high energy, which are scattered by interacting with the electrons in the molecules. X-rays with energy range of 5000-15000 eV are used for diffraction experiments. The diffraction pattern of real space molecules in a crystal is represented in a reciprocal space by Fourier transformation. The reciprocal lattice is a mathematical construct that describes diffraction events. The position of a reflection in reciprocal space is indicated by *h, k* and *l,* known as Miller indices. The diffraction of X-rays by a crystal can be quantified by Bragg's Law, which considers the diffracted X-rays as reflections from lattice planes.

The equation is defined as:

$$n\lambda = 2d_{hkl} \sin\theta$$

where, λ - wavelength

d_{hkl} - distance between two lattice planes

θ - scattering angle of X-ray beam

n - integer

The structure factor, F_{hkl}, determines the scattering of X-rays from a crystal in the direction hkl, and is a sum of all partial waves emitting from each and every atom of the unit cell. The intensity of scattered electromagnetic waves, I_{hkl}, is proportional to the square of the amplitude of the structure factor, and measured as diffraction spots. Fourier transformation of experimentally measured structure factor amplitudes along with phases generated from the phasing experiments produces electron density of the molecules. This electron density is represented in the form of a two-dimensional contour map or a three-dimensional grid, known as an electron density map. However, it is not possible to determine the phase of the diffracted X-ray directly by measuring its intensity. This means that during the measurement process, the phase information is lost, and this leads to the 'phase problem'. The information regarding the phases needs to be separately supplemented, in order to determine the structure of the protein.

2.3.2.3 Experimental Phasing by Molecular Replacement

There are several methods for solving the phase problem. These involve preparation of heavy atom derivatives, Se-Met labelling of protein or use of sulphur single-wavelength anomalous diffraction and collection of otherwise redundant data (Friedel pairs). However, another powerful, important and simple method to obtain initial phases is molecular replacement. This technique calculates phases from a known similar structure, generally having sequence identity of ~30% of more with the protein whose structure needs to be determined. The known protein is called the phasing model or the probe. In this method, the orientation of the probe in the unknown structure is determined by the rotation function, and the location of properly oriented probe in the unit cell is found by the translation function. Both the functions

are Patterson or maximum likelihood based. The phasing model is rotated and translated in the unit cell of the sought-after protein structure, until a best fit is obtained between the calculated diffraction data from the model and the observed data from the target structure.

2.3.2.4 Model Building and Refinement

After obtaining an initial model of the protein structure, it is fitted into the experimentally observed electron density, either manually or through automated programs. Model improvement and refinement is a cyclic process, and with every round, the phases and the electron density map improve, so allowing further building of the model and corrections. The model rebuilding-refinement cycles are continued until a good fit between observed and calculated structure factor amplitudes is obtained, which can be mathematically quantified by the residual index (R-value or R-factor). The lower the R-value, the better is the fit between the model and the experimental data.

$$R_f = \frac{\sum_{i=1}^{n} |F(i)_{obs} - F(i)_{cal}|}{\sum_{i=1}^{n} F(i)_{obs}}$$

where, R_f - residual index

$F(i)_{obs}$ - observed structure factor amplitude

$F(i)_{cal}$ - calculated structure factor amplitude

2.3.2.5 Structure Validation and Analysis

The structural model generated after sufficient rounds of rebuilding and refinement, is subjected to quality checks to determine its plausibility. The Ramachandran plot, which indicates how well the structure conforms to the stereo-chemical constraints, is very useful for validating the structure. A good model has the majority of residues tightly clustered in the favoured regions with minimum outliers, whereas a poor model has many outliers and no clustering. The structure is also checked to minimise errors concerning the bond angles and

lengths, chirality, dihedral angles, non-bonded atoms, planarity of aromatic rings and crystal packing.

2.3.3 Circular Dichoism

Circular dichroism (CD) is a method that is used to analyse the secondary structure of a protein, by utilizing differential absorption of left- and right-handed circularly polarised light produced by the peptide bonds in the polypeptide backbone. Different secondary structure elements have characteristic spectra in the far UV regions (Whitmore and Wallace, 2004, 2008).

2.3.4 Biochemical Characterisation using Coupled-Enzyme Colorimetric Assay

Production of trehalose-6-phosphate from NDP-glucose and glucose-6-phosphate, followed by the release of NDP, is estimated by measuring continuous oxidation of NADH by a coupled-enzyme colorimetric assay. This assay involves detection and quantification of the NDP released, when glucose is transferred from NDP–glucose to glucose-6-phosphate, by coupled conversion of phosphoenolpyruvate to pyruvate by pyruvate kinase, and of pyruvate to lactate by lactate dehydrogenase. In the final reaction, the oxidation of NADH is measured spectrophotometrically at 340 nm.

2.3.5 Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) is a quantitative technique that can measure the thermodynamic properties of protein-ligand interaction. It determines the heat evolved on binding of the ligand with its partner and directly measures the binding equilibrium. The energetics of binding complement the structural and functional studies, and provide a complete detail of the interaction forces (Freyer and Lewis, 2008; Milev, 2013; Pierce *et al*., 1999).

2.3.6 Thermal Shift Assay

The stability of protein is influenced by several factors, one of them being ligands that bind to them. The thermal stability can either increase or decrease upon ligand binding, and the change is monitored by a fluorescent dye (e.g. Sypro Orange). When the protein unfolds at a particular temperature, the dye binds to the exposed hydrophobic regions and the signal produced is quantified. The change in the melting temperature of the protein in the presence and absence of ligands or fragments is used to identify binding (Kranz and Schalk-Hihi, 2011; Niesen *et al*., 2007).

2.3.7 Fragment-Based Drug-Discovery

Fragment based drug-discovery (FBDD) is a powerful method which relies on structural biology and synthetic chemistry (Davies and Hyvönen, 2012). FBDD approach involves screening and validation of small molecules, of molecular weight < 300 Da, against the target protein, using an assortment of biophysical and structural techniques **(Figure 2.3)**. The hits obtained after screening are usually weakly binding to the target *in vitro* and have low affinity. These hits are then elaborated into lead compounds using fragment growing, linking or merging techniques to improve their potency (Scott *et al*., 2012). Fragment elaboration is an iterative process, directed by the structure of the target protein with fragments, as well as quantitative analysis of binding and biochemical assay to measure efficacy (Doak *et al*., 2016; Scott *et al*., 2012). The first FDA-approved drug developed by fragment based approach was Zelboraf (PLX4032) for treatment of late-stage melanoma (Bollag *et al*., 2010). FBDD has the following advantages:

a. a small fragment library allows exploration of a large region of chemical space.
b. the hit-rate of screening a fragment library is high,
c. optimised hits from fragment-screening have high atom-binding efficiency.

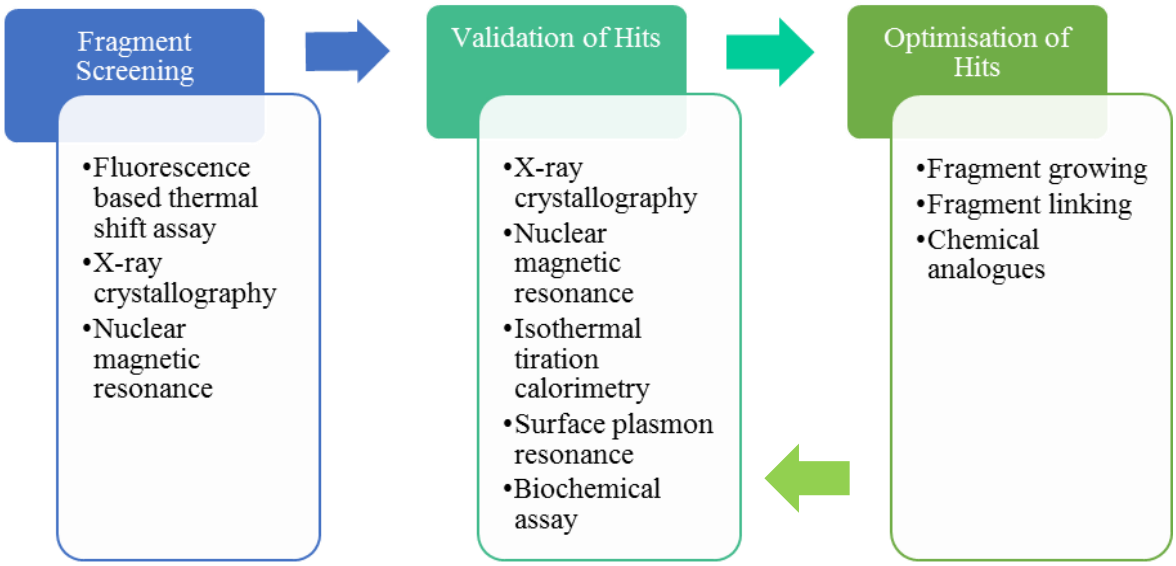


FIGURE 2.3 Typical process of a fragment-based drug-discovery approach.

(Figure adapted from Scott *et al*., 2012).

2.4 Methodology

2.4.1 Bioinformatic Analysis of the Primary Sequence, Sequence Alignment and Prediction of Secondary Structures

The amino-acid sequences of OtsA (Rv3490) from *M. tuberculosis* and of its orthologue in *M. smegmatis* (MSMEG_5892) were obtained from the TubercuList database (Lew *et al*., 2011) and the SmegmaList database (Kapopoulou *et al*., 2011) respectively, by searching with the gene name. Orthologous proteins in *M. thermoresistibile, E. coli, B. xenovorans, C. albicans, N. farcinica, S. cerevisiae, L. aerocolonigenes, S. hygroscopicus, D. melanogaster, C. felis, D. discoideum* and *A. thaliana* were identified by performing a protein-BLAST (blastp) (Altschul *et al*., 1990) search with *M. tuberculosis* protein sequence. Amino-acid sequences were obtained from the National Centre for Biotechnology Information (NCBI) database, and were aligned using Clustal Omega (Goujon *et al*., 2010; McWilliam *et al*., 2013; Sievers *et al*., 2014). The alignment was analysed and edited using Jalview (Waterhouse *et al*., 2009). Physicochemical properties of proteins were calculated using ProtParam (Gasteiger, 2003). Secondary structures and disordered regions of the

protein were predicted using PSIpred and DisoPred, respectively (Jones, 1999; Ward *et al*., 2004).

2.4.2 Identification of Structural Orthologues and Surface Conservation Analysis

A distance matrix search for structural homologues was performed using the DALI server (http://ekhidna.biocenter.helsinki.fi/dali_server/start) (Holm and Rosenstrom, 2010) to identify similar structures in the PDB. The conserved surface regions were identified using the ConSurf server (http://consurf.tau.ac.il/2016/) (Ashkenazy *et al*., 2010; Landau *et al*., 2005).

2.4.3 DNA, Bacterial Strains and Cloning of Mycobacterial *otsA* Genes

Genomic DNA from *M. tuberculosis* H37Rv was obtained from ATCC (ATCC25618D-2), and *M. thermoresistibile* (DSMZ 44167) genomic DNA was obtained from the Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ, Braunschweig, Germany). *E. coli* DH5α (Invitrogen) and BL21 (DE3) (New England Biolabs) strains were used for cloning and protein expression, respectively.

Dr. Vitor Mendes provided the glycerol BL21 cell stock for the expression clone of *M. thermoresistibile* OtsA pET-28a SUMO. The gene was cloned into BamHI and HindIII restriction sites in the pET-28a SUMO vector that has a cleavable, N-terminus SUMO tag and kanamycin resistance marker (obtained from Dr. Clifton Barry III's laboratory, National Institute of Health, Bethesda, USA).

otsA gene from *M. tuberculosis* genomic DNA was PCR amplified with the following primer pair:
Forward primer: 5'-ATACCATGGCTCCCTCGGGAGGCCAG-3'
(containing NcoI restriction site- underlined)
Reverse primer: 5'- ATTCTCGAGTTAGCCTTGGCCCCTCGGGTG-3'
(containing XhoI restriction site-underlined)

PCR was done with the KOD Hot Start DNA polymerase (Novagen), and the reaction was set up according to the manufacturer's protocol. DNA was denatured at 95°C for 2 min, followed by 30 cycles each comprising a 20 sec denaturation step at 95°C, 15 sec annealing step at 60°C, and 30 sec extension step at 70°C. A final extension step was done at 70°C for 5 min. The amplicon generated was cloned in NcoI and XhoI restriction sites of pHAT-4 vector (with a cleavble, N-terminus His_6 tag and ampicillin resistance marker) (Novagen). The gene and vector were digested with NcoI and XhoI restriction enzymes (ThermoScientific) and ligated using T4-DNA ligase (New England Biolabs) at room temperature for 10 min. The ligation products were transformed in *E. coli* DH5α competent cells, plated on LB agar-ampicillin plates, and incubated overnight at 37°C. Single colonies were randomly picked and inoculated in LB media (Melford) with 100 μg/ml ampicillin (Sigma), and grown overnight at 37°C, with constant shaking at 220 rpm. Plasmids were isolated and purified according to the manufacturer's protocol (ThermoScientific GeneJet Plasmid Miniprep Kit). The integrity of clones was confirmed by sequencing (DNA Sequencing Facility, Department of Biochemistry, University of Cambridge, UK), and plasmids were transformed into *E. coli* BL21 (DE3) strain.

2.4.4 Expression and Purification of Mycobacterial OtsA Proteins

2.4.4.1 Protein Expression Trial of *M. tuberculosis* OtsA

Transformed *E. coli* BL21 (DE3) cells of *M. tb* OtsA-His_6 were inoculated in 5 ml LB media containing 100 μg/ml ampicillin, and grown overnight with constant shaking (220 rpm) at 37°C. A small scale-up of the culture was done by inoculating 10 ml of fresh LB medium containing ampicillin, with 300 μl of the overnight culture. Cells were grown at 37°C till OD_{600} reached 0.8, then cooled to 20°C, induced with 0.5 mM IPTG (Melford) for protein expression and grown overnight (14-16 hours) at 20°C with constant shaking. The cells were harvested by centrifuging at 6200 x g, 4°C for 10 min. The pellet obtained was suspended in 1 ml resuspension buffer (50 mM Tris-HCl pH 7.5, 20 mM Imidazole, 500 mM NaCl and protease inhibitor cocktail (Roche)) and 10 μg/mL DNase (Sigma) was added. Cells were lysed with BugBuster protein extraction reagent (Novagen), at final concentration of 1 X, by incubating at room temperature for 10 min, and the lysate was centrifuged (20,000 x g, 4°C

for 10 min). The resulting supernatant was bound to the Ni-NTA agarose resin (Qiagen) packed in a 1 ml column (prepared in-house), with the help of a bench top centrifuge at 300 x g. The resin was washed twice with 500 µl wash buffer (50 mM Tris-HCl pH 7.5, 20 mM Imidazole, 500 mM NaCl) by centrifuging the column at 300 x g for 1 minute, and protein was eluted in 250 µl elution buffer (50 mM Tris-HCl pH 7.5, 500 mM Imidazole, 500 mM NaCl) by centrifuging the column at 950 x g for 1 min. The protein was analysed with SDS-PAGE. The protein's identity was confirmed by MALDI fingerprinting (PNAC facility, Department of Biochemistry, University of Cambridge, UK).

2.4.4.2 Expression and Purification of Full-Length Recombinant *M. thermoresistibile* OtsA

The recombinant orthologous protein of OtsA in *M. thermoresistibile* with a cleavable, N-terminus His_6-SUMO tag was purified.

Culture growth and expression: A loop of frozen BL21 cells of the expression clone was plated on an LB-agar-kanamycin plate and incubated overnight at 37°C. All the cells were collected from the plate and inoculated in 250 ml LB media with 30 µg/ml kanamycin. The culture was grown overnight at 37°C, with constant shaking at 220 rpm. The culture was scaled up to 6 litres by adding 30 ml of the overnight culture to each 1-litre of fresh LB media containing 30 µg/ml kanamycin. Cells were grown at 37°C till OD_{600} reached 0.75, cooled down to 20°C, induced with 0.5 mM IPTG for protein expression, and incubated overnight (14-16 hours) at 20°C with constant shaking.

Cell harvesting and lysis: Cells were harvested by centrifugation (4200 x g, 30 min, 4°C) during the late exponential growth phase and the pellet was re-suspended in buffer (20 mM Tris-HCl pH 7.5, 20 mM Imidazole, 500 mM NaCl, cocktail of protease inhibitor), and 10 µg/mL DNase-I and 5 mM $MgCl_2$ were added to the cell suspension. Cells were lysed by sonication, followed by centrifugation (30,000 x g, 40 min, 4°C) to remove the cell debris, and supernatant was collected (step-1). The pellet was suspended again in buffer (20 mM Tris-HCl pH 7.5, 20 mM Imidazole, 500 mM NaCl, 0.2% Triton X-100) and incubated on ice for 45 min, to extract any protein from the cellular fraction. The resuspension was centrifuged again (30,000 x g, 40 min, 4°C) to remove cell debris, and the supernatant was collected

(step-2). Supernatant from both step-1 and step-2 was pooled and filtered using 0.2 μm syringe filter.

Affinity purification: A pre-packed sepharose column (HiTrap IMAC FF, 5 ml, GE Healthcare) was charged with 0.1 M $NiSO_4$, according to the manufacturer's protocol, and equilibrated with purification buffers. Supernatant was bound to the column with the help of a peristaltic pump at room temperature. The column was washed for 30 CV with wash buffer (20 mM Tris-HCl pH 7.5, 20 mM Imidazole, 500 mM NaCl), and 30 ml protein was eluted using elution buffer (20 mM Tris-HCl pH 7.5, 500 mM Imidazole, 500 mM NaCl). The purity of the protein was determined by SDS-PAGE. Protein concentration was measured at 280 nm with NanoDrop spectrophotometer.

Cleavage of the tag and dialysis: The SUMO tag was subsequently cleaved with Ulp1 protease, after affinity purification. Ulp1 protease (see appendix-II for purification protocol of Ulp1) was added to the eluted protein at a ratio of 1 mg Ulp1 protease to 200 mg protein to cleave the His_6-SUMO tag. The protein-protease mixture was then dialysed in 20 mM Tris-HCl pH 7.5, 20 mM Imidazole and 500 mM NaCl buffer, using a 10 kDa MWCO dialysis membrane, and incubated overnight at 4°C.

Gel filtration: The overnight-digested protein was concentrated by centrifuging (4000 x g, 30 min, 4°C) in 10 kDa MWCO concentrators. The protein was loaded onto a Superdex-200 column (HiLoad 16/600 Superdex 200 pg, GE Healthcare), pre-equilibrated with 500 mM NaCl and 20 mM Tris-HCl pH 7.5. Purification was carried out at room temperature. The protein was eluted in 1 ml fractions and the purity of fractions was determined by SDS-PAGE. The purest fractions were pooled and concentrated by centrifuging (4000 x g, 30 min, 4°C) in 10 kDa MWCO concentrators. The concentration of NaCl was reduced from 500 mM to 350 mM with 20 mM Tris-HCl pH 7.5 buffer. The purity of the protein was assessed with SDS-PAGE and the final concentration was determined spectrophotometrically. The protein's identity was confirmed by MALDI fingerprinting (PNAC facility, Department of Biochemistry, University of Cambridge, UK).

2.4.5 Circular Dichroism

0.5 mg/ml *M. thermoresistibile* OtsA, dialysed in a buffer of 20 mM sodium phosphate pH 7.5 and 330 mM NaF, was placed in quartz cuvette with 1 mm path-length. The circular dichroism (CD) spectra were recorded on Aviv model 410 spectropolarimeter. A wavelength spectrum was recorded from 260 nm to 190 nm, at 1 nm intervals and 3 sec averaging time. A control run was performed with the buffer. For the protein sample, the raw data from the three measurements were averaged, corrected for the buffer signal, and smoothed. A temperature melt spectrum was recorded at a constant wavelength of 222 nm from 25˚C to 95˚C, at 1˚C interval and 3 sec averaging time. The CD signal is reported in theta machine units in millidegrees.

2.4.6 Biochemical Characterisation of Recombinant *M. thermoresistibile* OtsA

The reactions were done at 37 ºC and a typical 100 µl enzyme reaction contained 20 mM Tris-HCl pH 7.5, 60 mM NaCl, 0.3 mM NADH, 2.5 mM phosphoenolpyruvate, 0.7 units of pyruvate kinase/lactate dehydrogenase, 2.5 µM OtsA and varying concentrations of glucose-6-phosphate and NDP-glucose to determine the kinetic parameters. All reagents used were obtained from Sigma-Aldrich.

The substrate specificity of the enzyme was examined using ADP-glucose, GDP-glucose and UDP-glucose as glucose donors (at a fixed concentration of 2.5 mM); and glucose-6-phosphate, glucose-1-phosphate and fructose-6-phosphate as glucose acceptors (at fixed concentration of 5 mM). The effects of cations and pH were determined by the above mentioned coupled-enzyme colorimetric assay, with 100 µl mix containing appropriate divalent cation and buffer. The effects of cations were tested by incubating reaction mixtures with chloride salts of Mg^{2+}, Mn^{2+}, Zn^{2+}, Ca^{2+}, Fe^{2+} and Co^{2+} (5 mM and 10mM). The pH profile was determined in 20 mM acetate buffer at pH 5.0- 5.5, 20 mM MES buffer at pH 6.0-6.5, 20 mM Tris-HCl buffer at pH 7.0-9.0, and 20 mM CAPS buffer at pH 9.5-10.0. A temperature-dependent profile was not studied, as *M. thermoresistibile* is a thermophile and its proteins are too thermostable (Edwards *et al*., 2012). The effects of trehalose and trehalose-6-phosphate as feedback inhibitors of *M. thermoresistibile* OtsA were tested in the

conditions mentioned above with fixed concentrations of glucose-6-phosphate (5 mM) and ADP-glucose (2.5 mM).

All the reactions were done in triplicate in a UV-1800 Shimadzu UV spectrophotometer and data were analysed with Prism 5 (GraphPad Software) and Microsoft Excel.

2.4.7 Crystallisation of *M. thermoresistibile* OtsA and Data Collection

2.4.7.1 Crystallisation of Apo *M. thermoresistibile* OtsA protein and Data Collection

Initial crystallisation screening was performed for purified *M. thermoresistibile* OtsA protein using the sitting-drop vapour-diffusion method at 18˚C. The screens used were: classics (Qiagen), cryo (Rigaku), PEGs-I, JCSG+ (Molecular Dimensions) and Wizard I & II (Rigaku). The protein (0.3 µl) at a concentration of 10 mg/ml was mixed with the well solution in two different protein: precipitant ratios of 1:1 and 1:2 using the Phoenix robot (Art Robbins Instruments). Initial crystals were obtained in several conditions, of which, a condition from the PEGs-I crystallisation screen was optimised further. This condition is composed of 40 % PEG-200 and 0.1 M Tris-HCl pH 8.5, with 1:1 protein: precipitant ratio. It was further optimised using a 24-well Linbro plate and hanging-drop vapour-diffusion method. The grid was prepared manually with different concentrations of PEG-200 (30 %, 35 %, 40 %, 45 %, 50 %, and 55 %) and 0.1 M Tris-HCl at varying pH (7.5, 8.0, 8.5, and 9.0). The condition obtained from this grid (35 % PEG-200, 0.1 M Tris-HCl pH 9.0, and protein: precipitant =1:1) was optimised using an additive screen HT (Hampton Research), and LDAO was found to be the best additive. The additive containing condition was further optimised, and the final optimised condition contained 40 % PEG-200, 0.3 % LDAO, with 2:1 protein: precipitant ratio. The crystal condition contained PEG-200, at a high enough concentration for it to act as a cryoprotectant. Thus, the crystals were directly flash frozen in liquid nitrogen for data collection. An X-ray diffraction dataset of the crystal was collected at I02 beamline of Diamond Light Source synchrotron (DLS, Oxford, UK).

2.4.7.2 Soaking and Co-Crystallisation of *M. thermoresistibile* OtsA with Substrates and Products

The ligand-bound structures of *M. thermoresistibile* OtsA were obtained by soaking the apo-crystals with ligands (ADP-glucose, UDP-glucose, GDP-glucose, trehalose and trehalose-6-phosphate) in the optimised crystal condition using the hanging-drop vapour-diffusion method. Crystals were then transferred to a 2 µl drop of reservoir solution containing 10 mM of ligand and incubated for 18-20 hours. The crystals were directly flash frozen in liquid nitrogen for data collection.

The ternary complex of *M. thermoresistibile* OtsA with ADP and glucose-6-phosphate (G6P) was obtained by co-crystallising the protein with 5 mM ADP and 5 mM G6P. An initial screening was done using the screens: classics, JCSG+, Wizard I & II, and Wizard III & IV. Best crystals were obtained from the JCSG+ screen, in the condition containing 0.3 M magnesium formate dehydrate and 0.1 M Bis-Tris-HCl pH 5.5. Crystals were cryoprotected by soaking them briefly in a solution containing the mother liquor and 25% ethylene glycol, and were flash frozen in liquid nitrogen for data collection.

The binary complex of *M. thermoresistibile* OtsA-ADP was obtained by co-crystallising the protein with 5mM ADP. Initial screening was done using the screens: classics (Qiagen), JCSG+ (Molecular dimensions), Wizard I&II (Rigaku), and Wizard III&IV (Rigaku). The initial crystals were obtained from the Wizard I&II screen, in the condition containing 1.2 M ammonium sulphate and 0.1 M HEPES pH 7.5. This was further optimised by using an additive screen HT, and 2, 5-hexanediol was found to be the best additive. Crystals were cryoprotected by soaking them briefly in a solution containing the mother liquor and 25% ethylene glycol, and were flash frozen in liquid nitrogen for data collection.

The X-ray diffraction datasets of the soaked-crystals and co-crystals were collected at I03, I04, and I04-1 beamlines of Diamond Light Source synchrotron, Oxford, UK. The interactions between ligands and protein were calculated using Arpeggio (http://structure.bioc.cam.ac.uk/arpeggio).

2.4.8 Structure Solution and Refinement of OtsA Datasets

The diffraction datasets were processed using MOSFLM (Battye *et al*., 2011) and reduced using AIMLESS (Evans and Murshudov, 2013), both tools from the CCP4 suite (Winn *et al*., 2011). The resulting integrated and scaled file was used for molecular replacement to determine initial phases using PHASER (McCoy *et al*., 2007) from the PHENIX software package (Adams *et al*., 2010). The structure of *E. coli* OtsA (PDB entry: 1UQU) was used as the phasing model. The model was built building with Coot (Emsley *et al*., 2010), and refinement was performed in REFMAC 5 (Murshudov *et al*., 1997) of the CCP4 suite or PHENIX (Adams *et al*., 2010), as well as manually. The structure was validated using Coot, and REFMAC 5 or PHENIX. The figures were prepared with Pymol (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.) and UCSF Chimera (Pettersen *et al*., 2004).

2.4.9 Isothermal Titration Calorimetry

The thermodynamics of the binding between *M. thermoresistibile* OtsA and ligands (ADP-glucose, UDP-glucose, GDP-glucose, glucose-6-phosphate) was characterised using a Microcal ITC200 titration calorimeter at 25ºC. Protein at concentration of 100 µM in 20 mM Tris-HCl pH 7.5, 330 mM NaCl was used for all titrations. 1 mM ligands were injected in 2 µl aliquots, and 18 injections with 120 sec spacing were done. Data were analysed by fitting a simple single-site model using Origin software (Microcal).

2.4.10 Fluorescence-Based Thermal Shift Assay

The initial screening of a fragment library (kindly provided by Prof. Abell's group, Department of Chemistry, University of Cambridge) against *M. thermoresistibile* OtsA was carried out using a fluorescence-based thermal shift assay. The assay was performed in a 96-well plate format, using a CFX Connect real-time PCR machine (Bio-Rad). Each 25 µl well contained 20 µM protein, 100 mM HEPES pH 7.5, 350 mM NaCl, 5X Sypro Orange

fluorescence dye (Invitrogen) and 5 mM fragments. Fragments were dissolved in 100% DMSO. 5 mM ADP-glucose and 5% DMSO were used as the positive and negative controls respectively.

The fragment library was also screened in the presence of ADP. The reaction mixture was the same as mentioned above, but contained 5 mM ADP in addition. For this screening, 5 mM ADP-glucose and 5 mM ADP were used as the positive and negative controls respectively.

The thermal scan was recorded for each sample from 25°C to 95°C, at 0.5°C increments. The fluorescence was measured every 30 sec. The best binders were selected on the basis of ΔTm and the shape of the melting curve. The data were analysed using an in-house Microsoft Office Excel macro.

2.4.11 Soaking and Co-crystallisation with Fragment Hits

To obtain the fragment-bound structures of *M. thermoresistibile* OtsA, the apo-crystals produced using the hanging-drop vapour-diffusion method were soaked in the optimised crystal condition containing the fragments. Crystals were then transferred to a 2 µl drop of reservoir solution containing different concentrations of ligand (20 mM, 50 mM, 100 mM dissolved in 100% DMSO) and incubated for 14-16 hours. The crystals were directly flash frozen in liquid nitrogen for data collection.

To obtain strutures bound with the fragments hits obtained from the thermal shift screening of the enzyme along with ADP, *M. thermoresistibile* OtsA was co-crystallised with 5 mM ADP and each of 5 the fragments at a concentration of 5 mM. Initial screening was done using the screens: wizard I&II, wizard III&IV and JCSG+. Crystals were obtained in several conditions, but the following conditions were common for all the fragments tested:

1. JCSG+ screen, solution C9 (0.1 M Sodium/potassium phosphate pH 6.2, 25 % 1,2-Propandiol 10 % Glycerol)
2. Wizard I&II screen, solution A2 (10% 2-propanol, 100 mM HEPES pH 7.5, 200 mM Sodium chloride)
3. Wizard I&II screen, solution B1 (1.26 mM Ammonium sulfate, 100 mM Sodium cacodylate pH 6.5)

4. Wizard III&IV screen, solution E3 (100 mM Bis-TRIS pH 5.5, 200 mM Ammonium sulfate)

Crystals of protein with fragment(s) and ADP were also obtained in the condition where the protein was co-crystallised with ADP and G6P (condition containing 0.3 M magnesium formate dehydrate and 0.1 M Bis-TRIS pH 5.5). Different cryoprotectants were used (30% glycerol, 25% ethylene glycol, 30% PEG-400, 25% MPD, 1:1 Paratone-N: Paraffin). Crystals were cryoprotected by soaking them briefly in a solution containing the mother liquor and the cryo-protectant, and were flash frozen in liquid nitrogen for data collection.

All the datasets were collected in beamlines I03 and I04 of Diamond Light Synchrotron (Oxford, UK).

2.4.12 Inhibition Assays

The inhibition assay was performed, as described in section 2.4.6, with varying concentrations of the fragments. The value of IC_{50} was calculated from the dose-response curve for each fragment.

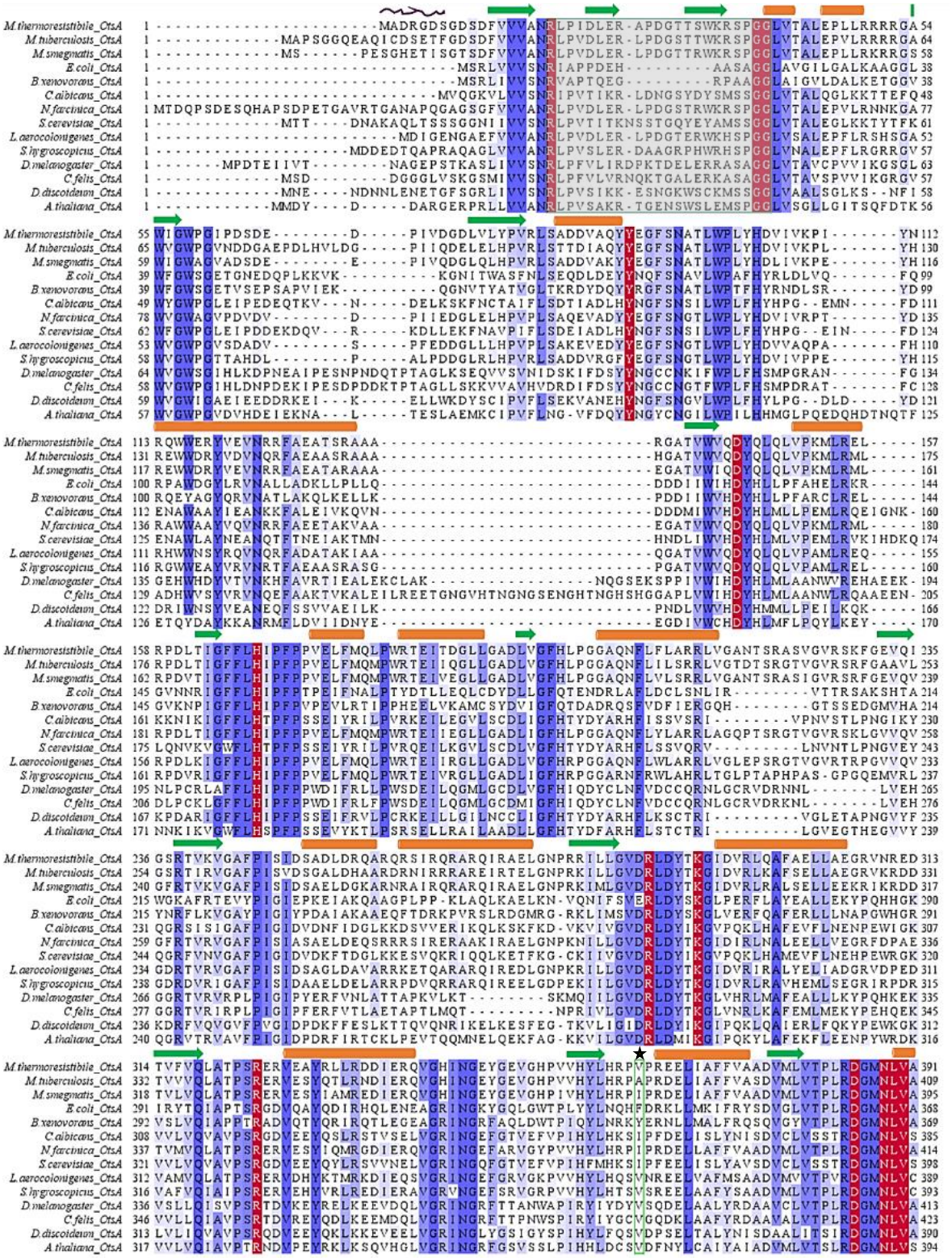
2.5 Results

2.5.1 Alignment and Analysis of OtsA Sequences

OtsA is found in diverse species, such as bacteria, fungi, insects, plants. Based on the BLAST analysis of *M. tuberculosis* OtsA protein sequence, it exhibits homology to OtsA proteins in different prokaryotes and eukaryotes, including *M. thermoresistibile* (80% identity), *M. smegmatis* (79% identity), *E. coli* (34% identity), *B. xenovorans* (35% identity), *C. albicans* (37% identity), *N. farcinica* (77% identity), *S. cerevisiae* (37% identity), *L. aerocolonigenes* (69% identity), *S. hygroscopicus* (64% identity), *D. melanogaster* (35% identity), *C. felis* (36% identity), *D. discoideum* (37% identity) and *A. thaliana* (36% identity). The sequences were aligned using Clustal Omega, and the predicted secondary structure elements of *M. thermoresistibile* OtsA were mapped onto the alignment.

The residues of the catalytic site of *E. coli* OtsA, determined from the crystal structure of the protein complexed with UDP and glucose-6-phosphate (PDB ID: 1GZ5) (Gibson *et al*., 2002a), were marked. The alignment shows that in OtsA enzyme from all these species, despite sharing limited sequence similarity, the residues of the catalytic site of *E. coli* OtsA are conserved, with the exception of Phe339 **(Figure 2.4)**. In *E. coli* OtsA, this residue confers the specificity for the uracil moiety (Gibson *et al*., 2002a).

The loop between Arg9 and Gly22 of *E. coli* OtsA is catalytically important, and is the shortest among the other OtsA sequences (Gibson *et al*., 2002a, 2004). The enzyme from *B. xenovorans* also has a short loop like that of the *E. coli* enzyme. In other organisms, including mycobacteria, there is an insertion of 8-9 residues in this loop.


M.thermoresistibile_OtsA
M.tuberculosis_OtsA
M.smegmatis_OtsA
E.coli_OtsA
B.xenovorans_OtsA
C.albicans_OtsA
N.farcinica_OtsA
S.cerevisiae_OtsA
L.aerocolonigenes_OtsA
S.lyrgoscopicus_OtsA
D.melanogaster_OtsA
C.felis_OtsA
D.discoideum_OtsA
A.thaliana_OtsA

(cont.)

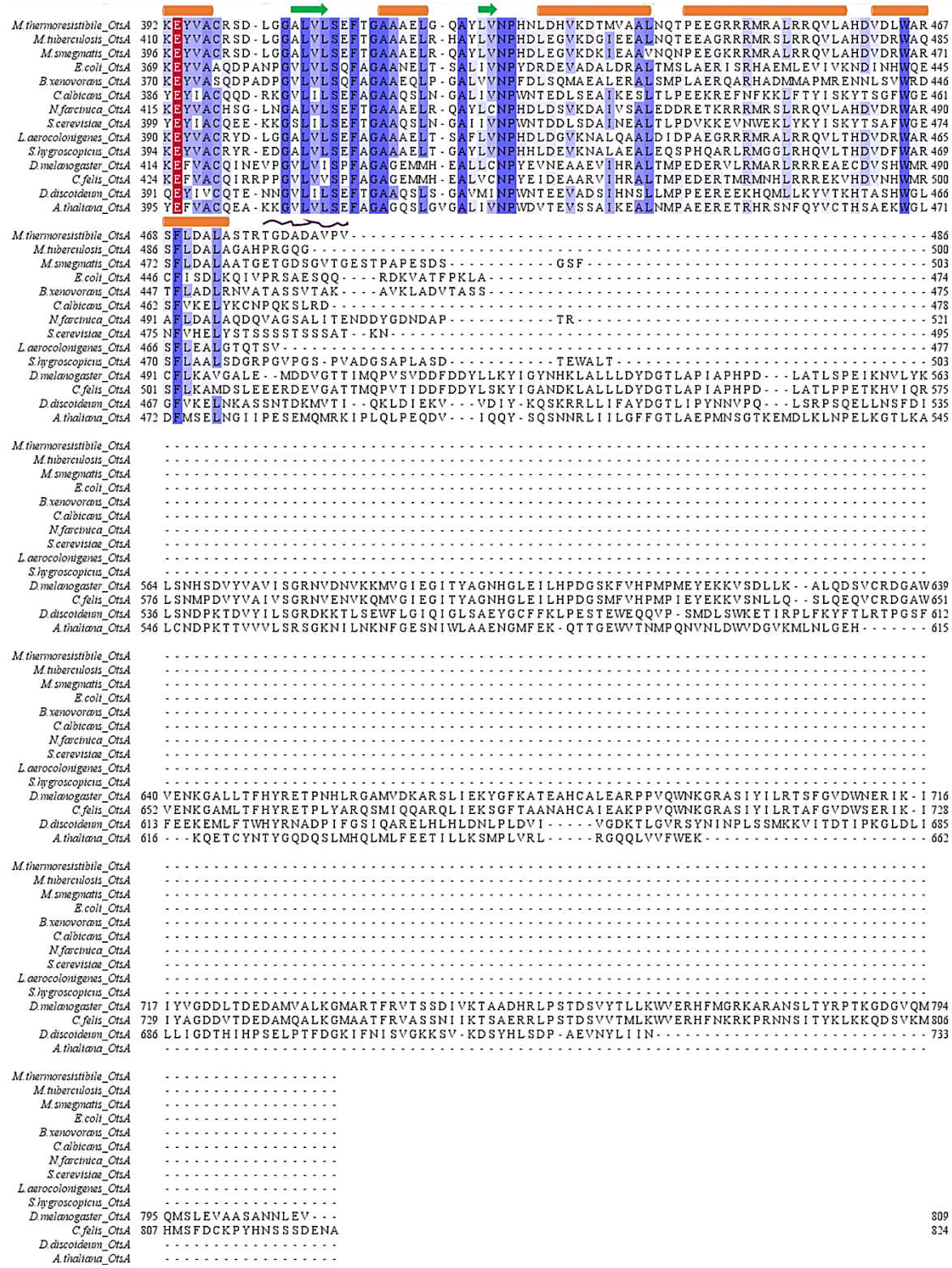


FIGURE 2.4 Sequence alignment of OtsA sequences from different organisms.

Amino-acid sequence alignment of OtsA in bacteria, fungi, insects and plants. Sequence alignment is annotated with predicted secondary structural features of *M. thermoresistibile* OtsA (orange tube: α- helix; green arrow: β-strand; purple curvy line: disordered region). The catalytic site loop is shaded grey. Conserved residues of the substrate-binding pocket of *E. coli* OtsA (Gibson *et al*., 2002a) are highlighted maroon. Other residues are shaded according to the degree of conservation. The unconserved residue, Phe339, of the catalytic pocket of *E. coli* OtsA and the corresponding residues in other species in the alignment are marked with a black star.

2.5.2 Expression Trial of *M. tuberculosis* OtsA

An expression trial of N-terminal His_6-tagged *M. tuberculosis* OtsA showed that the protein was expressed in the insoluble fraction. A strong band of His_6-tagged *M. tuberculosis* OtsA observed around 58 kDa was seen in the pellet fraction **(Figure 2.5)**.

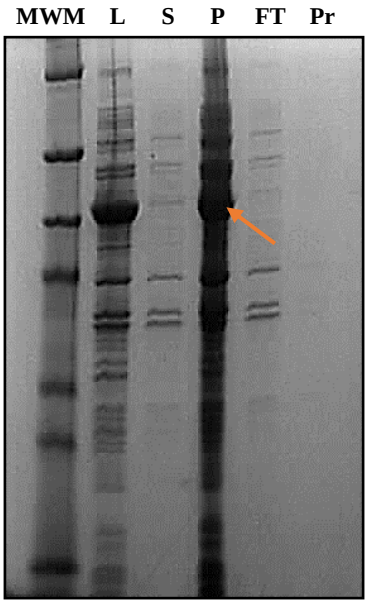


FIGURE 2.5 Expression trial of His_6-tagged *M. tuberculosis* OtsA.

Expression trial of His_6-tagged protein on Ni-sepharose column. MW of the tagged protein- 58 kDa. Orange arrow: tagged protein in the pellet fraction.
(MWM: molecular weight marker, L: cellular lysate, S: supernatant, P: pellet, FT: flow through, Pr: tagged protein).

The results of expression trials done by Dr. Vitor Mendes indicated that SUMO-tagged *M. thermoresistibile* OtsA was present in the soluble fraction, whereas the SUMO-tagged *M. tuberculosis* protein was mostly insoluble and found in inclusion bodies (Dr. Vitor Mendes, personal communication). Thus, the SUMO-tagged orthologous protein from *M. thermoresistibile* was expressed and purified.

2.5.3 Expression and Purification of the Full-Length Recombinant *M. thermoresistibile* OtsA

The full-length orthologous protein was expressed with a cleavable N-terminus SUMO tag in *E. coli.* The initial affinity purification indicated that protein was expressed in both soluble and pellet fractions. Further, the protein from the cellular fraction was extracted by a very mild detergent treatment step during affinity purification, which increased the overall yield by almost 30%. The SUMO tag cleaved by the Ulp1 protease was separated, and the protein was further purified by gel filtration. High-levels of protein, purified to homogeneity, were produced. MALDI fingerprinting confirmed that a strong band observed at 58 kDa was of the full-length *M. thermoresistibile* OtsA, and a low molecular weight band was the degradation product of the full-length protein **(Figure 2.6)**. The trace from the gel-filtration run indicated that the tag-less protein behaved as a tetramer in the solution. The purified orthologous protein was used for subsequent structural studies, biochemical and biophysical assays.

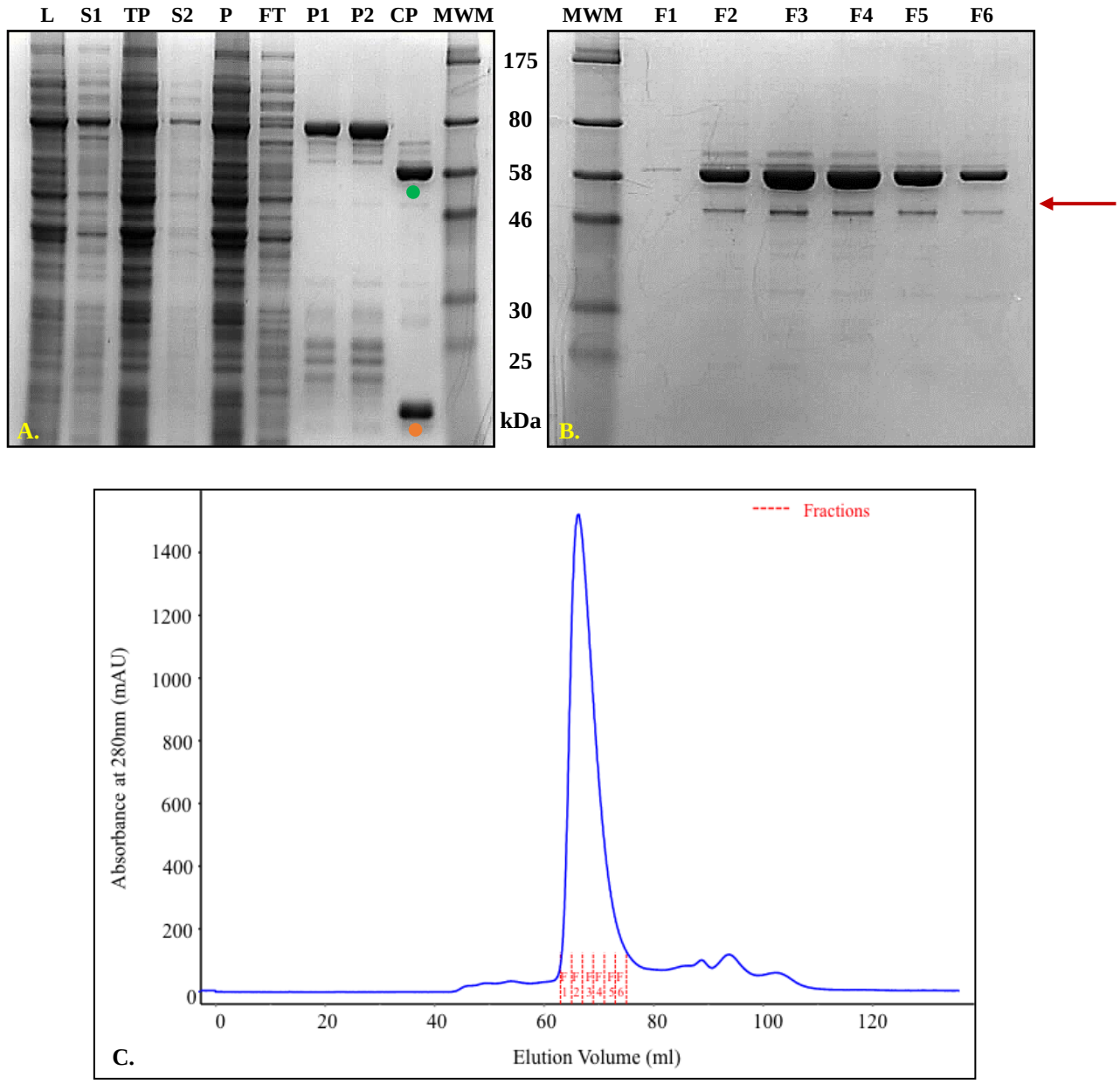


FIGURE 2.6 Expression and purification of *M. thermoresistibile* OtsA.

A. Metal affinity purification of SUMO tagged protein on Ni-sepharose column and cleavage of SUMO tag by UlpI protease. MW of the protein- 54.5 kDa, MW of SUMO tag-13.8 kDa, MW of the tagged protein-68.3 kDa. Green circle: tag-less protein after Ulp1 cleavage, orange circle: SUMO tag.
(MWM: molecular weight marker, S1: supernatant from soluble fraction, TP: Triton X-100 treated pellet, S2: supernatant after detergent treatment, P: pellet, FT: flow through, P1 to P2: tagged protein, CP: protein cleaved with Ulp1 protease)

B. Gel filtration of the cleaved protein on Superdex-200 column. Maroon arrow represents the degradation band of the full-length protein.
(MWM: molecular weight marker, F1 to F6: tag-less protein fractions)

C. The gel filtration profile. The protein eluted at around 65 ml.

2.5.4 Analysis of the Secondary Structure and Thermodynamic Stability by Circular Dichroism

A circular dichroism (CD) experiment was conducted on *M. thermoresistibile* OtsA to identify any detrimental effect of the mild detergent treatment on the protein structure. The far-UV CD spectrum showed that the protein was folded, and was a mix of α-helices and β-strands **(Figure 2.7 A)**. The thermal stability of *M. thermoresistibile* OtsA was examined by measuring ellipticity over the temperature range of 25-95˚C. This was important in determining whether the protein had any unfolded intermediates due to the mild detergent treatment during purification. The protein displayed a monophasic, steep melting curve, indicating absence of any intermediates. The unfolding of the protein occurred over a narrow temperature range between 64˚C and 73˚C **(Figure 2.7 B)**.

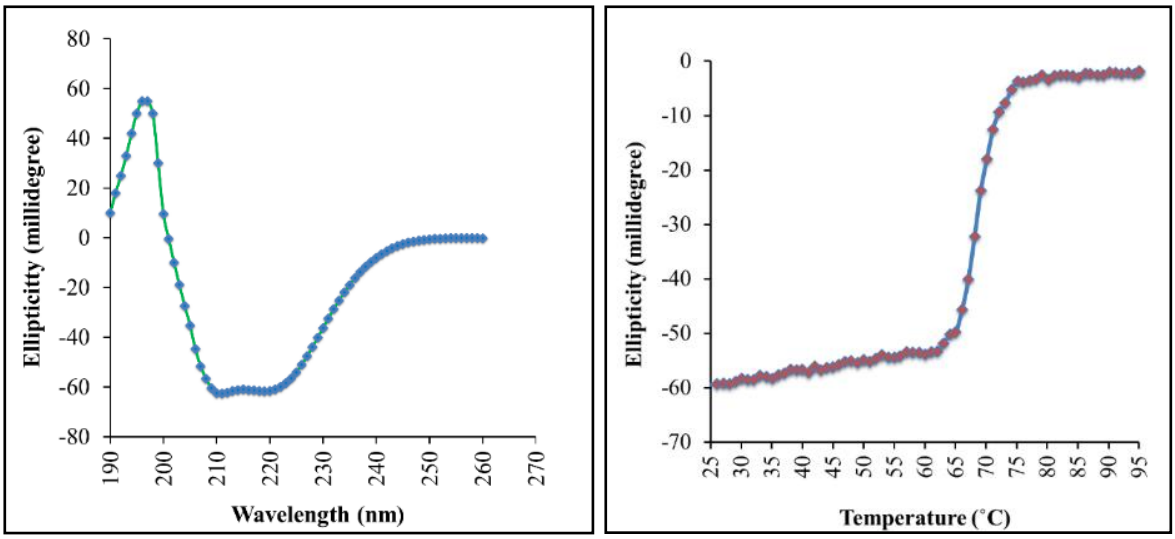


FIGURE 2.7 The circular dichroism spectra of full-length *M. thermoresistibile* OtsA.

A. Wavelength scan indicates that the protein is folded with a combination of α-helices and β-strands.

B. Thermal profile shows a narrow range for the melting temperature.

2.5.5 Properties of *M. thermoresistibile* OtsA, Substrate-Preference and Regulation of Enzyme Activity by Feedback-Inhibition

M. tuberculosis OtsA and *M. thermoresistibile* OtsA belong to glucosyltransferase family-20 of the CAZy classification (Carbohydrate Active Enzymes database, www.cazy.org) (Campbell *et al*., 1997; Lombard *et al*., 2014). The kinetic parameters for *M. thermoresistible* OtsA were determined with K_m values of 0.17 ± 0.03 mM for ADP-glucose, and of 0.23 ± 0.04 mM for glucose-6-phosphate **(Table 2.1, Figure 2.8)**.

SUBSTRATE	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m ($M^{-1}s^{-1}$)
ADP-Glucose	0.17 ± 0.03	1.75 ± 0.08	10300 ± 2600
Glucose-6P	0.23 ± 0.04	4.22 ± 0.23	18300 ± 5700

TABLE 2.1 Kinetic parameters for *M. thermoresistibile* OtsA obtained in this study. (K_m: Michaelis Menton constant, k_{cat}: turnover number, k_{cat}/K_m: catalytic efficiency)

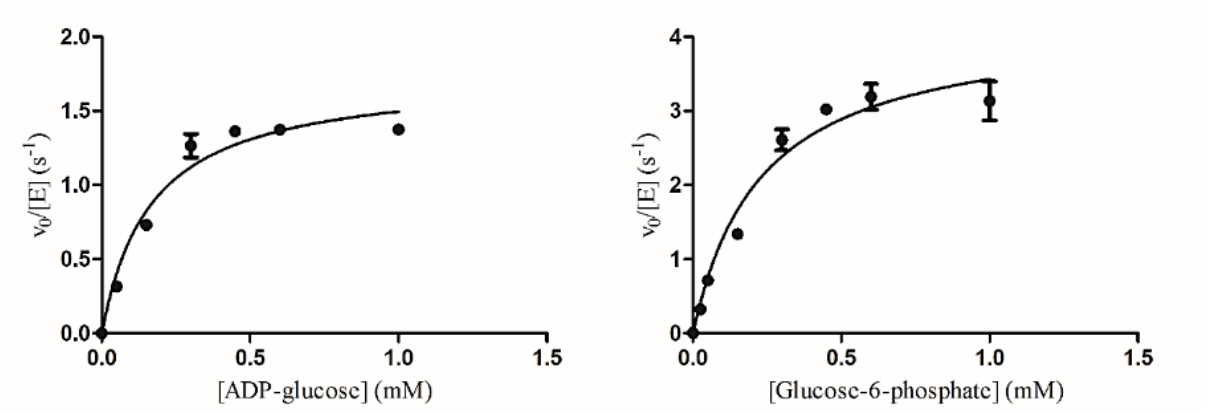


FIGURE 2.8 Kinetics of *M. thermoresistibile* OtsA with ADP-glucose and glucose-6-phosphate.

The activity of recombinant *M. thermoresistibile* OtsA was not dependent on divalent cations. There was no substantial difference in the activity of the enzyme upon adding Mg^{2+}, Mn^{2+}, Zn^{2+}, Ca^{2+}, Fe^{2+} and Co^{2+}. However, the activity was slightly higher with Mg^{2+} and Mn^{2+} than other tested divalent cations **(Figure 2.9 A)**. The effect of pH was also tested on the activity of the enzyme. *M. thermoresistibile* OtsA was active between the pH 5.0 and 10.0 at 37˚C, exhibiting maximum activity between pH 7.0-8.5 **(Figure 2.9 B)**.

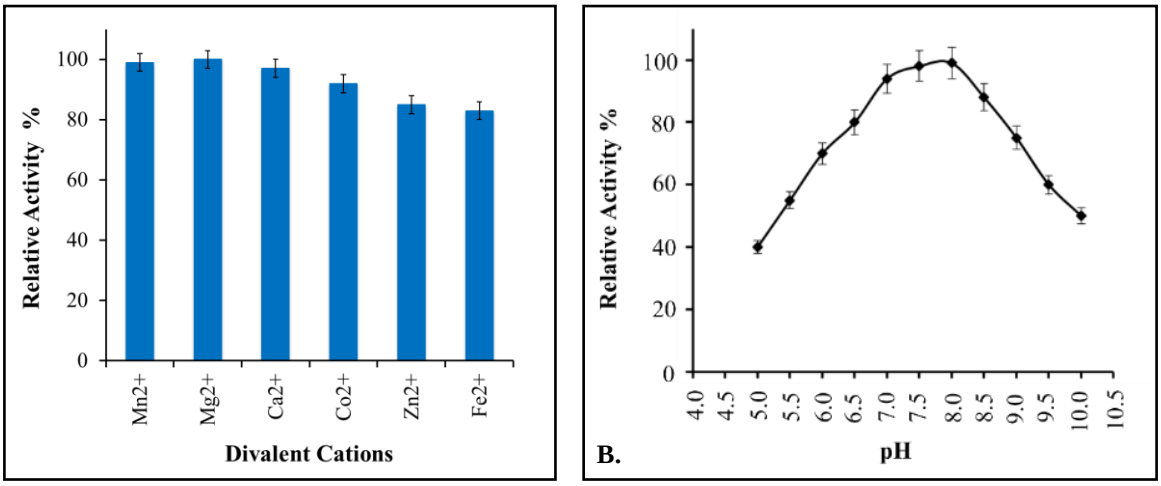


FIGURE 2.9 Effect of divalent ions (A) and pH (B) on the activity of *M. thermoresistibile* OtsA.

M. thermoresistibile OtsA used ADP-glucose, GDP-glucose and UDP-glucose as glucose donors with decreasing preference, and had strict specificity towards glucose-6-phosphate as the glucose acceptor **(Figure 2.10)**, in conformity with what has been reported previously for *M. tuberculosis* OtsA (Asención Diez *et al*., 2015; Pan *et al*., 2002). The ITC data further established the preference of ADP-glucose over other glucose donors. Binding of ADP-glucose was determined in the tested conditions with an observable K_d of 21.88 µM ± 2.00 µM. The complexation of ADP-glucose with *M. thermoresistibile* OtsA was endothermic (positive peaks in the ITC output) **(Figure 2.11)**. The association was characterised by an unfavorable enthalpy change (ΔH°= +6002 cal/mol) and a positive entropy change (ΔS= +41.5 cal/mol/deg) (values of thermodynamic properties are shown in an inset in the figure 2.11). Other glucose donors, including GDP-glucose and UDP-glucose, and the glucose acceptor G6P, showed no observable heat of binding in the tested conditions.

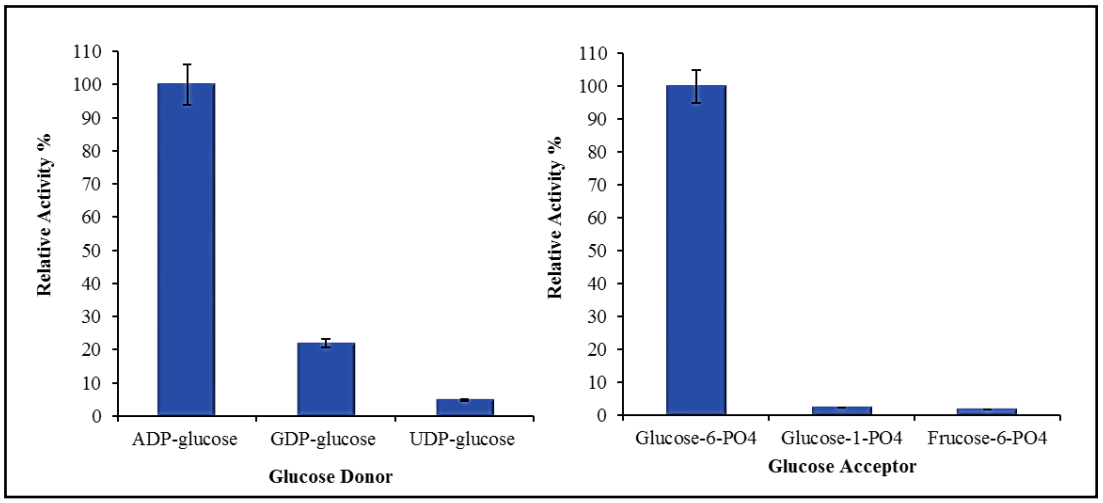


FIGURE 2.10 Substrate preference of *M. thermoresistibile* OtsA.

The enzyme showed highest preference for ADP- glucose as its substrate for glucose donor (left graph), and had strict preference for glucose-6-phosphate as the acceptor (right graph).

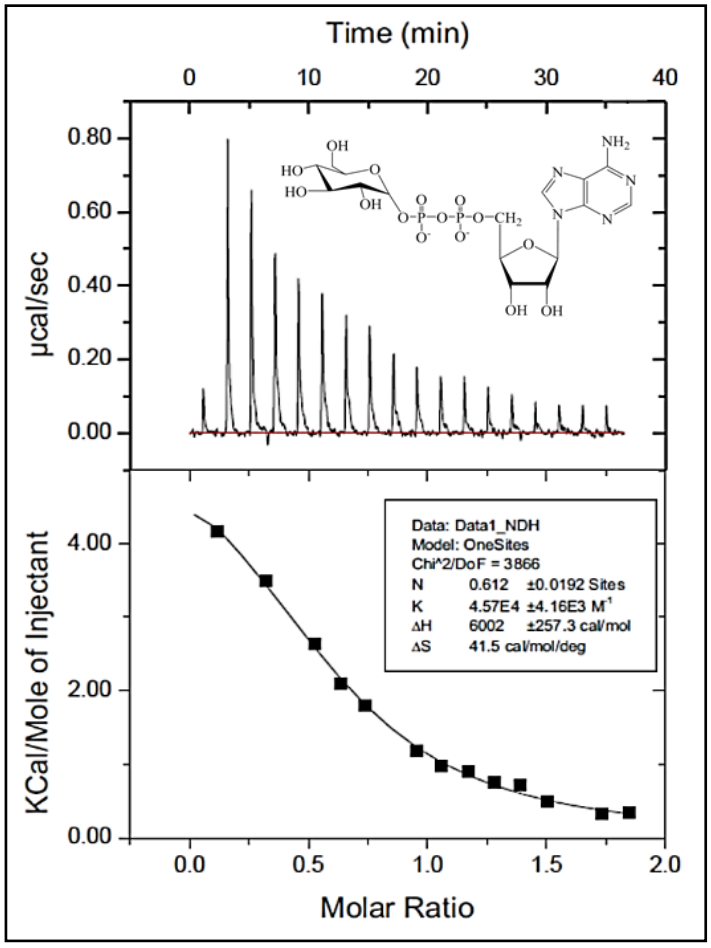


FIGURE 2.11 Isothermal titration calorimetry profile of *M. thermoresistibile* OtsA with ADP-glucose.

It was also shown that the enzyme activity was regulated by both trehalose-6-phosphate, which is the reaction product, and trehalose, the pathway product, thus indicating feedback inhibition of the enzyme activity. Trehalose-6-phosphate inhibited the enzyme activity by approximately 40% at 2 mM, whereas trehalose had a more prominent inhibitory effect with almost 50% inhibition at 1 mM **(Figure 2.12)**.

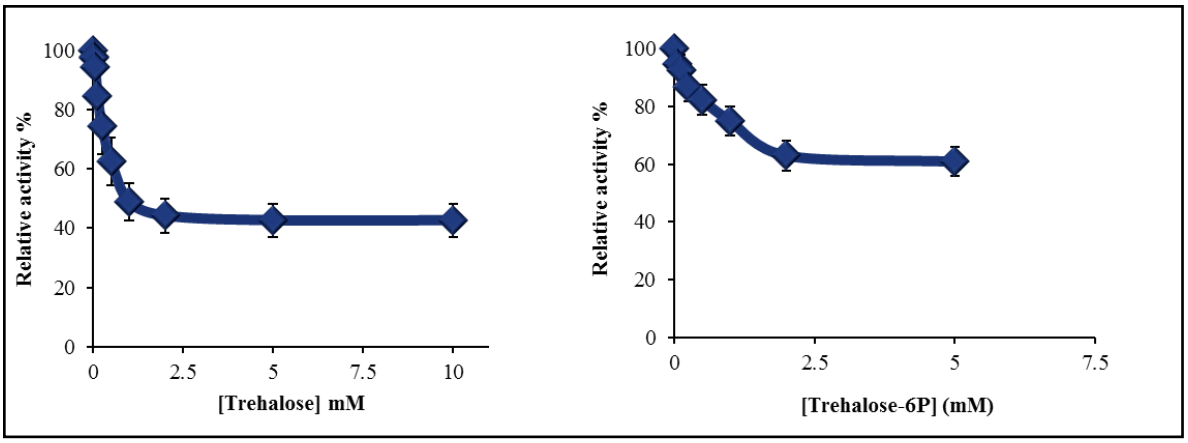


FIGURE 2.12 Feedback inhibition of *M. thermoresistibile* OtsA by trehalose and trehalose-6-phosphate.

Enzyme activity was inhibited by 50% at 1 mM trehalose (left graph) and by 40% at 2 mM trehalose-6-phosphate (right graph).

2.5.6 Overall Structure of Apo, Full-Length *M. thermoresistibile* OtsA

The crystals of *M. thermoresistibile* OtsA diffracted to 2.18 Å resolution **(Figure 2.13)** and belonged to $F4_132$ space group, with two protomers per asymmetric unit **(Figure 2.14 A)**. Superposition of the two protomers with BATON showed that the RMSD between them was 0.182 Å. The unit cell parameters were **a** = **b** = **c** = 337.5 Å; **α** = **β** = **ɣ** = 90˚. The Matthew's coefficient was calculated to be 2.45, and the protein crystals had 49.79 % solvent content. The final protein structure was modelled for residues 7-23, 33-323 and 331-477 as the density was not visible for the remaining residues. The data collection and refinement statistics for apo-enzyme crystals are detailed in **Table 2.2**. Each protomer of *M. thermoresistibile* OtsA consists of two domains having the classic Rossmann-fold, and the catalytic site is situated between the two domains **(Figure 2.14 B)**. The N-terminal domain consists of residues 1-247 and 462-486, which constitute a core of 7 parallel β-strands with an antiparallel β-strand on

either side, and 8 α-helixes, of which one helix is made of the endmost residues of the C-terminal domain. The C-terminal domain, spanning the residues 248-461, is made of 6 parallel β-stands intercalated with 9 α-helixes. The last helix of this domain has a kink and extends towards the N-terminal domain **(Figure 2.15 A and B)**. The apo-protein (and in bound state with many ligands, described in section 2.5.7) adopts an open conformation. The N-terminal domain has the highest values for atomic temperature factors, which indicats that this domain is more dynamic **(Figure 2.16)**. This is concomitant with the movement of α1 helix during the conformational changes of the enzyme essential for its catalytic activity (described in detail in section 2.5.8).

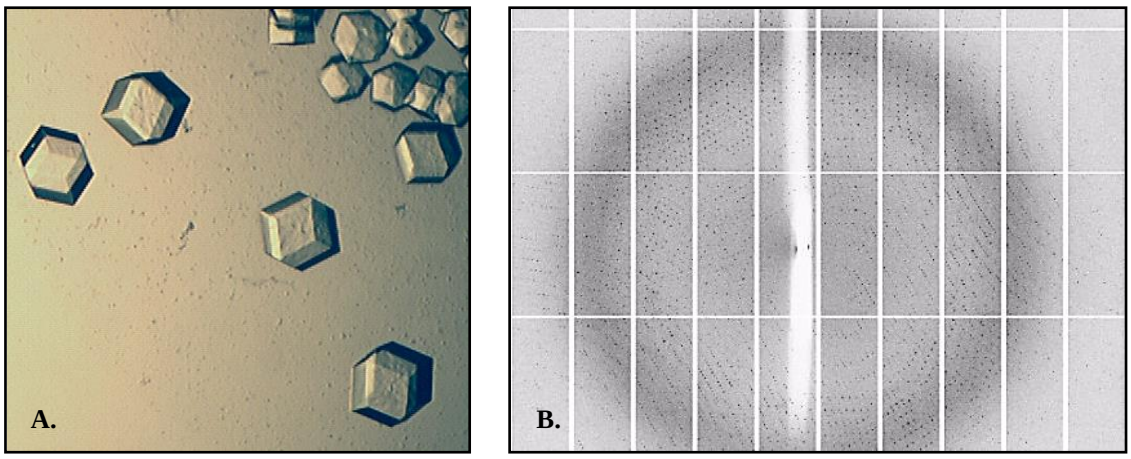


FIGURE 2.13 Apo-enzyme crystals and diffraction pattern of full-length *M. thermoresistibile* OtsA

A. Cubic crystals of *M. thermoresistibile* OtsA.
B. X-ray diffraction pattern of the crystals.

	Apo *M. thermoresistibile* OtsA
Data collection*	
Beamline	Diamond Light Source, I02
Space group	$F4_132$
Cell parameters:	
a/ b/ c (Å)	337.47/ 337.47/ 337.47
α/ β/ ɣ (˚)	90/90/90
Resolution range (Å)	50.88 – 2.18 (2.24 – 2.18)
No. of reflections:	
Observed	1980805 (67062)
Unique	85478 (6222)
Completeness (%)	100 (99.9)
R_{meas}	0.109 (0.742)
I/ σ (I)	26.4 (4.1)
Refinement	
Refinement program	REFMAC 5
Resolution	50.88 – 2.18
No. of reflections	84195
R_{work}/R_{free} (%)	18.96/20.63
R.M.S. Deviations:	
Bonds [Å]	0.007
Angles [˚]	1.056
Ramachandran Plot:	
Favored [%]	98.3
Outliers [%]	0
B- factors (Å):	
Protein atoms	33.8
Solvent	42.6

Table 2.2 Data collection and refinement parameters of apo *M. thermoresistibile* OtsA structure.

*Parameters in parenthesis are for the highest resolution shell

DLS: Diamond Light Source, Oxford, UK

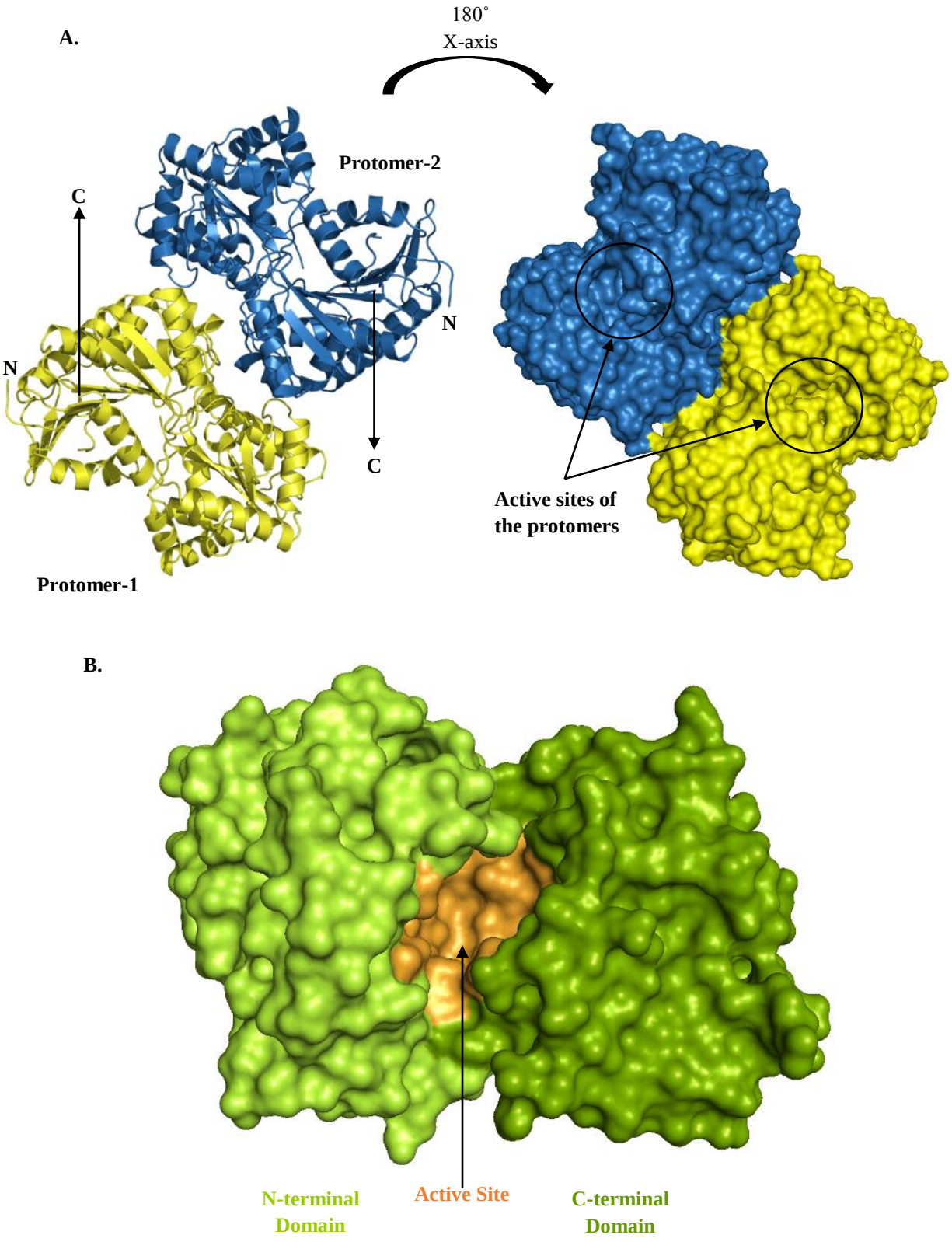


FIGURE 2.14 The crystal structure of apo, full-length *M. thermoresistibile* OtsA.

A. Cartoon representation of the homodimer in the asymmetric unit. The dimer is rotated 180˚ and represented in molecular surface to show the relative positioning of active sites of the protomers in the dimer.

B. Surface representation of the protein structure, showing two-domain architecture of the protein. The catalytic pocket is coloured in orange. Only one protomer is shown.

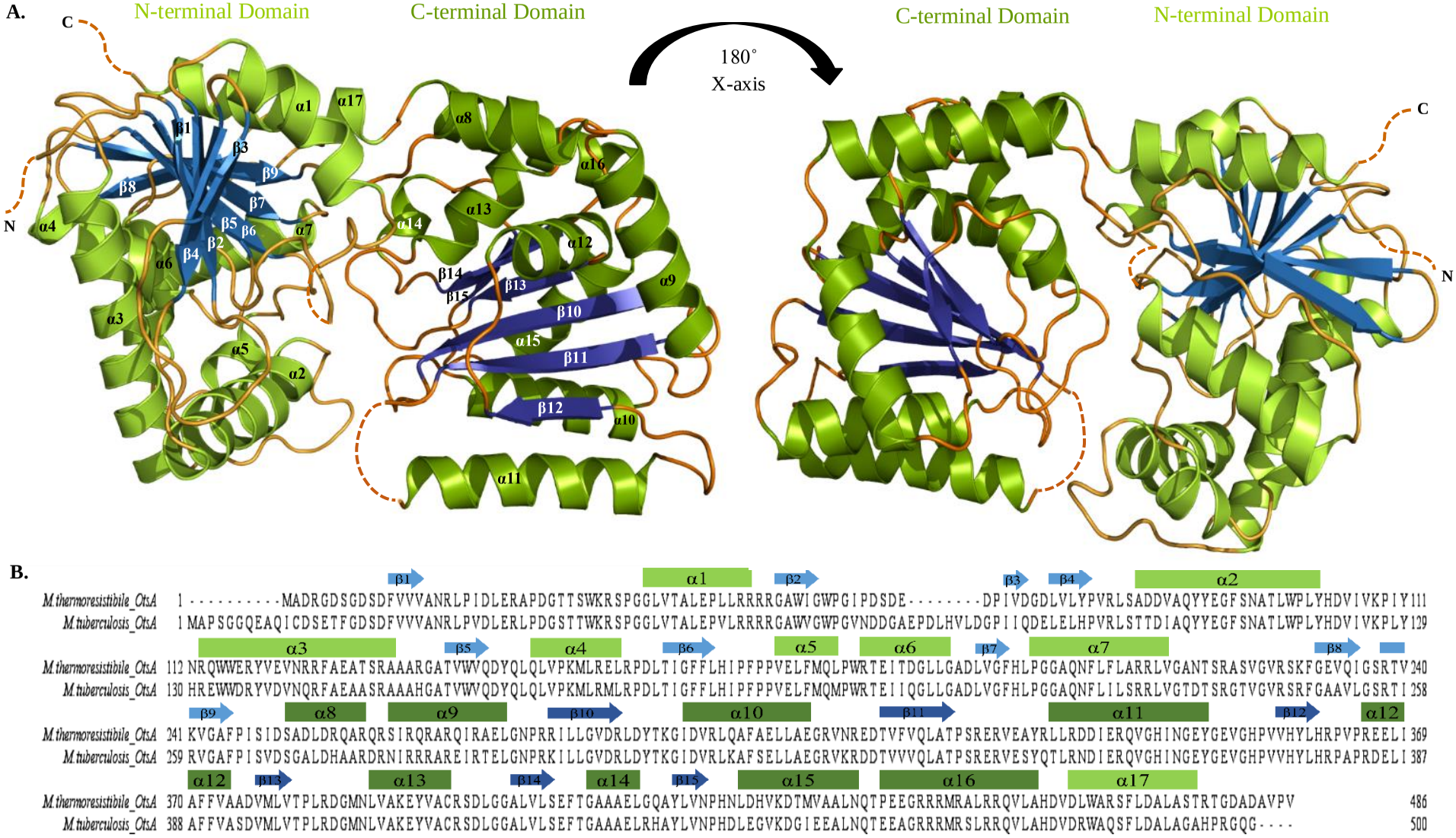


FIGURE 2.15 The crystal structure of apo, full-length *M. thermoresistibile* OtsA.

A. Cartoon representation of the overall protein structure, showing two Rossmann-fold domains. Dashed lines indicate residues for which the electron density was not visible. Only one protomer is shown. The α-helices and β-strands are numbered in both the domains in sequential order from the N-terminus.

B. Protein sequence of full-length *M. thermoresistibile* OtsA annotated with structural elements.

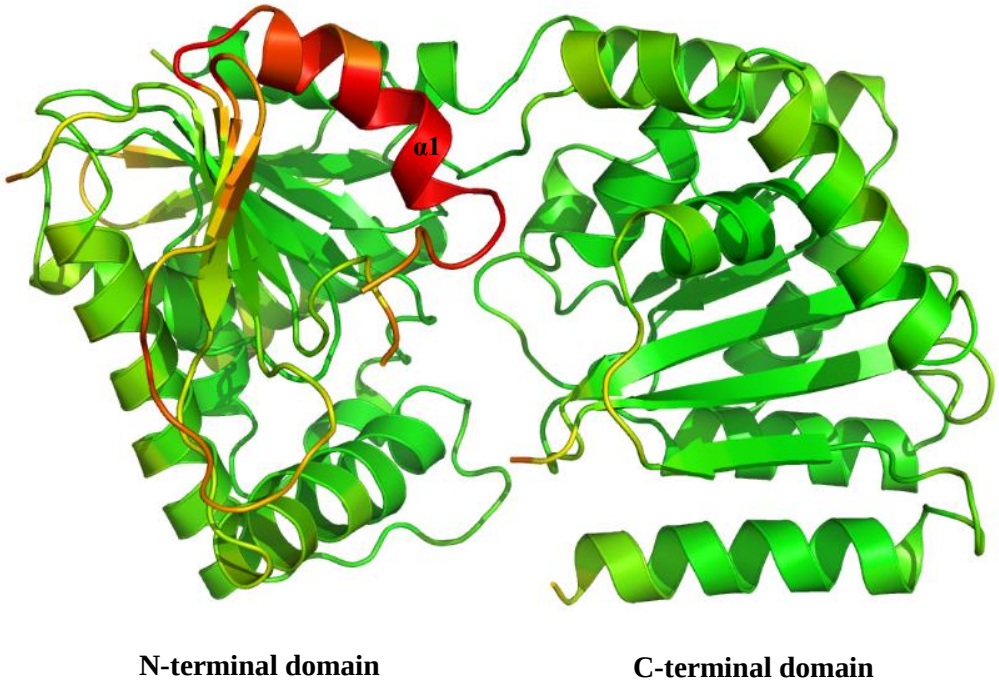


FIGURE 2.16 Crystallographic B-factor distribution in apo, full-length *M. thermoresistibile* OtsA.

Cartoon representation of the structure colored by the B-factor (Debye-Waller or temperature factor). B-factor values are shown with colors ranging from red (high fluctuations) to green (low fluctuations). N-terminal domain has the highest B-factor value, particularly the α1 helix, indicating high flexibilty of this domain. Only one protomer is shown.

The surface conservation analysis of *M. thermoresistibile* OtsA showed that the catalytic core, at the interface of N- and C-terminal domains of the enzyme, was conserved. This was expected as it confers to the trehalose-6-phosphate synthase activity, and is under stronger evolutionary pressure. However, a few residues, Leu319, Thr 321, Arg361 and Val 363, of the catalytic binding site of the enzyme, particularly the donor-site (described in section 2.5.7) do not show strong evolutionary conservation, and are of particular interest **(Figure 2.17)**. This probably accounts for the relaxed preference that OtsAs exhibit for the nitrogenous bases of the nucleotide donor-substrate.

Apart from the catalytic core, there were conserved residues at the dimerisation and tetramerisation interfaces, detailed in the following section.

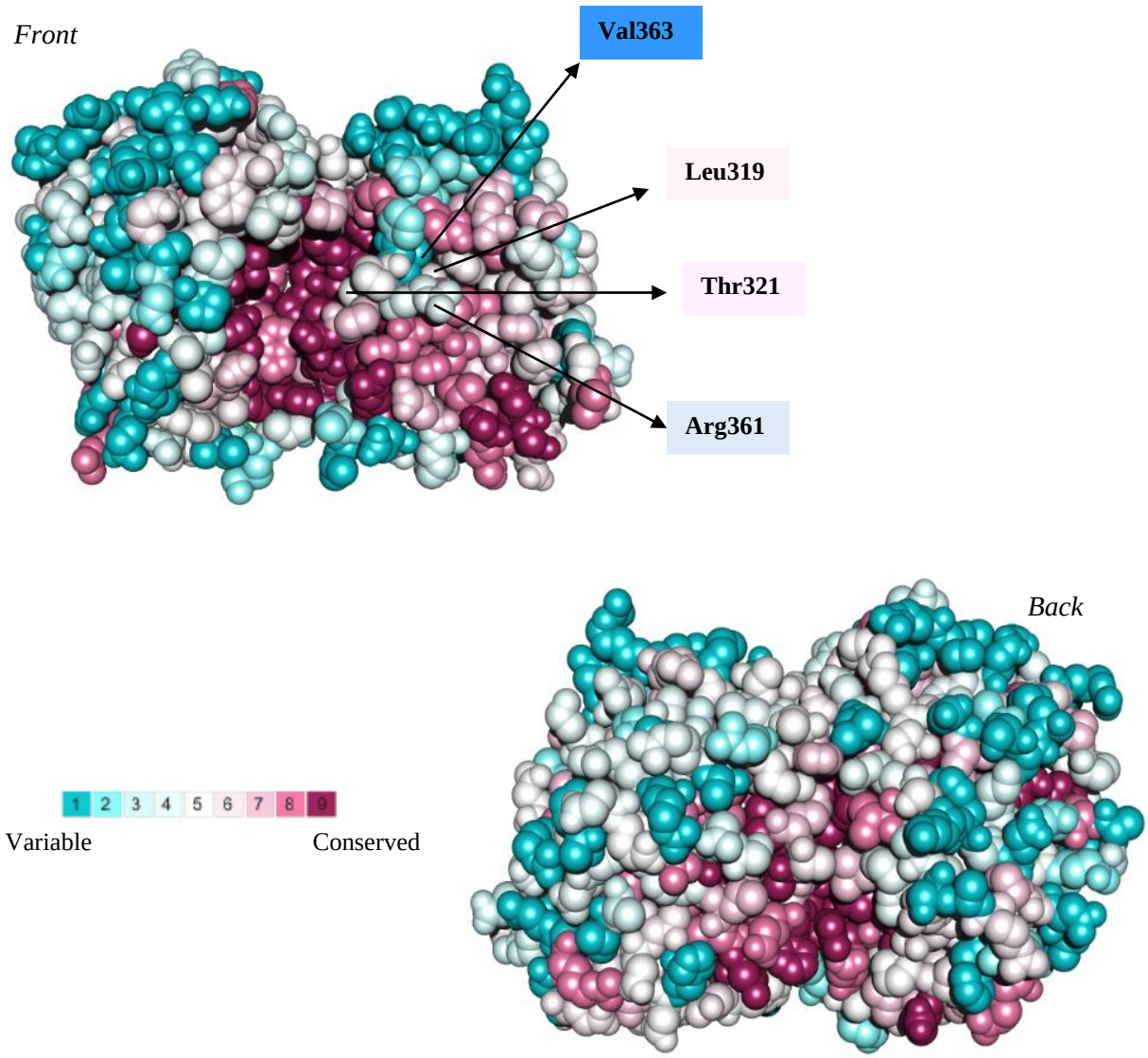


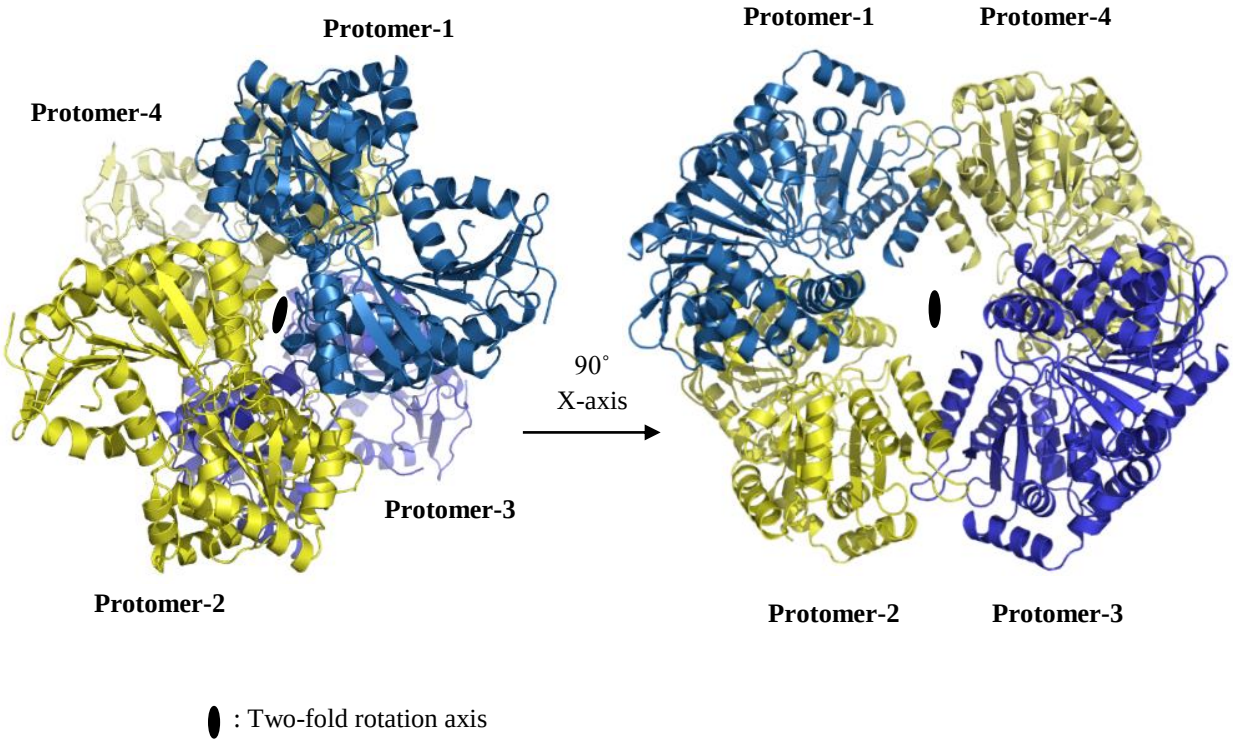
FIGURE 2.17 Conservation analysis of *M. thermoresistibile* OtsA.

The sequence conservation pattern of *M. thermoresistibile* OtsA**,** obtained using the Consurf server, shown in two views related by a 180° rotation. The conserved and variable residues, presented as space-filled model, are colored from the least conserved (turquoise) through intermediately conserved positions (white) to the most conserved residues (burgundy). The unconserved residues at the donor-subtrated site of the catalytic pocket are labelled.

Quaternary Structure of M. thermoresistibile OtsA

M. thermoresistibile OtsA forms a tetrameric assembly in the crystal-structure **(Figure 2.18 A)**. The asymmetric unit contains two protomers that are related by a two-fold axis to form a homo-dimer **(Figure 2.18 A)**. The two protomers interact through a large dimerisation interface, with residues residing mainly on α5 and α7, and loops between α2- α3, α5- α6, β6-α5, α7-β8, β13-α13, β14-α14 and β15-α15 **(Figure 2.18 B and C)**. Many interfacial residues are conserved, notably, Phe171, Glu175, Pro181, Arg213, Arg384, Phe410, Asn 424 and Pro425. Further, a dimer is formed between each such dimer related by a two-fold axis to form a homo-tetrameric assembly **(Figure 2.18 A)**. The two dimers contact through α11 and the loop between α11-β12 **(Figure 2.18 D and E)**. This interface contains a few invariant residues: Arg334, Gly342, Asn345 and Gly346.

A.



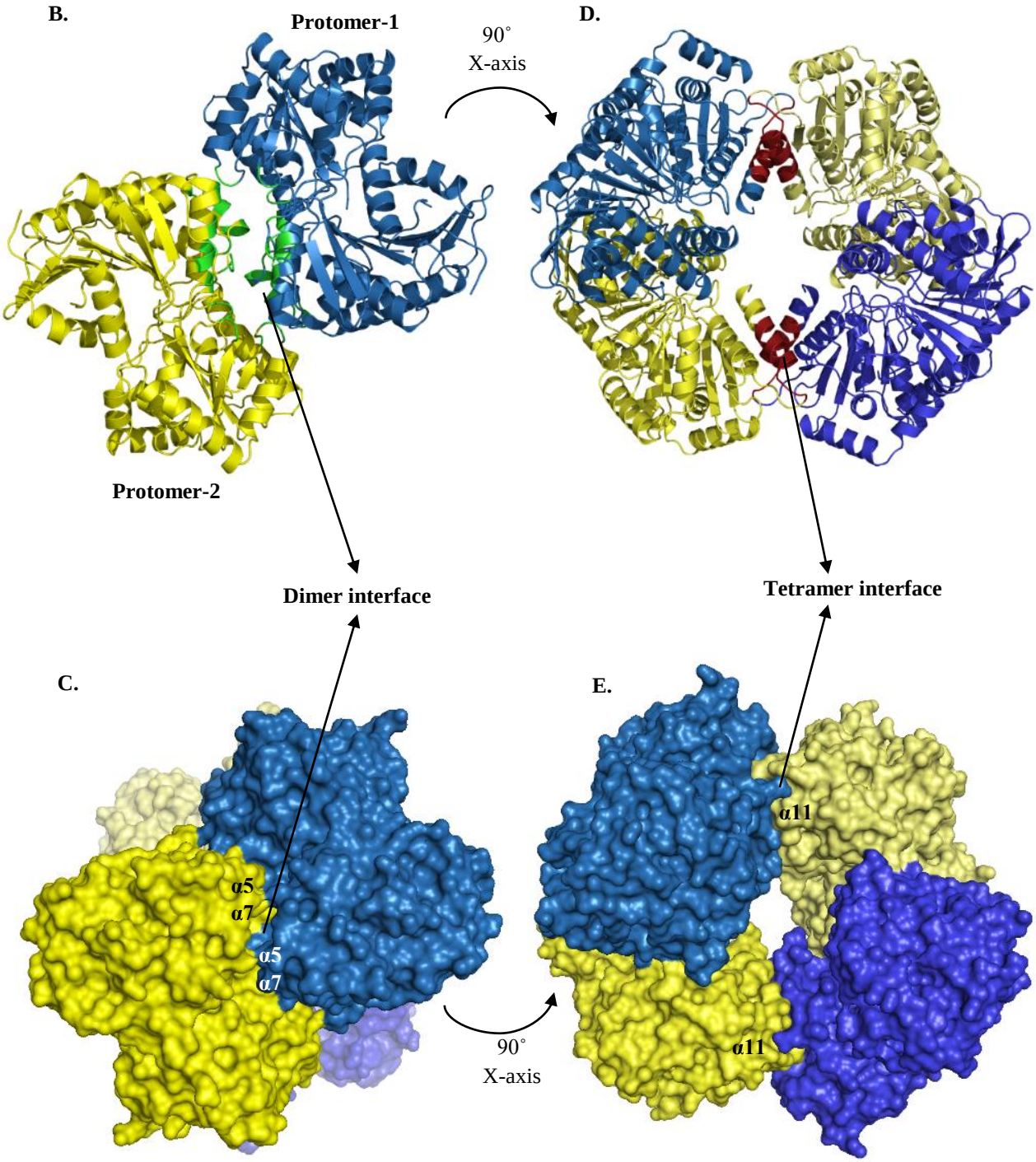


FIGURE 2.18 Tetrameric assembly of apo *M. thermoresistibile* OtsA.

- **A.** Cartoon representation of *M. thermoresistibile* OtsA dimer in the asymmetric unit along the two-fold rotation axis. 90˚ rotated view displays the entire tetrameric assembly of the protein, along the twofold rotation axis.
- **B.** Side view of *M. thermoresistibile* OtsA tetramer in cartoon representation, showing the dimerisation interface between the protomers, highlighted in green. Same dimerisation interface is observed between the protomers 3 and 4.
- **C.** Molecular surface representation of B.
- **D.** Front view of *M. thermoresistibile* OtsA tetramer in cartoon representation, displaying the tetramer interface between the homo-dimers, highlighted in maroon.
- **E.** Molecular surface representation of D.

2.5.7 The Donor Site of *M. thermoresistibile* OtsA: Binding of Glucose-Donor Substrates

Ligand-bound structures of *M. thermoresistibile* OtsA were determined for glucose-donor substrates: ADP-glucose, GDP-glucose and UDP-glucose, by soaking the apo-crystals with these ligands. For clarity, only one protomer of the protein-ligand complex is shown in the figures. The structure with inactive nucleotide-substrate, ADP, was obtained by co-crystallisation, and had one protomer in the asymmetric unit. The data collection and refinement statistics for ligand-bound crystals are detailed in **Table 2.3**.

	ADP-glucose	**GDP-glucose**	**UDP-glucose**	**ADP**
Data collection*				
Beamline	DLS, I04	DLS, I04	DLS, I04	DLS, I04-1
Space group	$F4_132$	$F4_132$	$F4_132$	$I4_122$
Cell parameters:				
a/ b/ c (Å)	337.55/337.55/337.55	337.03/337.03/337.03	337.10/337.10/337.10	129/129/209.7
α/ β/ ɣ (˚)	90/90/90	90/90/90	90/90/90	90/90/90
Resolution range (Å)	76.38-2.19 (2.24-2.19)	56.97-2.29 (2.54-2.29)	40.88-4.94 (5.11-4.94)	109.91-1.88 (1.93-1.88)
No. of reflections:				
Observed	2769099 (89607)	2633027 (95731)	2377933 (91426)	888152 (514347)
Unique	109260 (6624)	111391 (8113)	120145 (11638)	71808 (5243)
Completeness (%)	99.9 (99.8)	100 (100)	100 (99.9)	99.9 (99.8)
R_{meas}	0.121 (0.843)	0.069 (0.679)	0.085 (0.691)	0.083 (2.304)
I/ σ (I)	36.2 (4.8)	37.7 (4.0)	23.7 (3.1)	20.3 (2.1)
Refinement				
Refinement program	REFMAC 5	REFMAC 5	REFMAC 5	PHENIX
Resolution	76.38-2.19	56.97-2.29	40.88-4.94	48.74-1.88
No. of reflections	108882	111227	119745	71544
R_{work}/R_{free}	17.72/19.51	17.88/21.03	19.41/22.76	21.89/23.44
R.M.S. Deviations:				
Bonds [Å]	0.009	0.008	0.012	0.009
Angles [˚]	1.049	1.053	1.087	1.037
Ramachandran Plot:				
Favored [%]	98.4	97.7	98.1	98.5
Outliers [%]	0	0	0	0

Table 2.3 Data collection and refinement parameters of glucose-donor substrates bound *M. thermoresistibile* OtsA structures.

*Parameters in the parenthesis are for highest resolution shell
DLS: Diamond Light Source, Oxford, UK

The prime interactions of the glucose-donor substrates are with the C-terminal domain. The side chain contacts are made with Arg286, Lys291, Arg365, Asp385 and Glu393; and the backbone contacts are made with the absolutely conserved residues Gly386, Met387, Asn388 and Leu389. The interaction of the donor substrate with the N-terminal domain is mediated by His168 and His199. The hydroxyl-groups of the ribose-ring coordinate with the amide group of Arg365 and carbonyl groups of Glu393. The phosphate groups interact with the amide groups of Arg286, Lys291 and Leu389. The glucose-ring interacts with amide groups of His168, His199, Gly386, Met387 and Asn388; and carbonyl group of Asp385 **(Figure 2.19 and Figure 2.20)**. The protein adopts an open conformation in all the observed structures with the donor-groups.

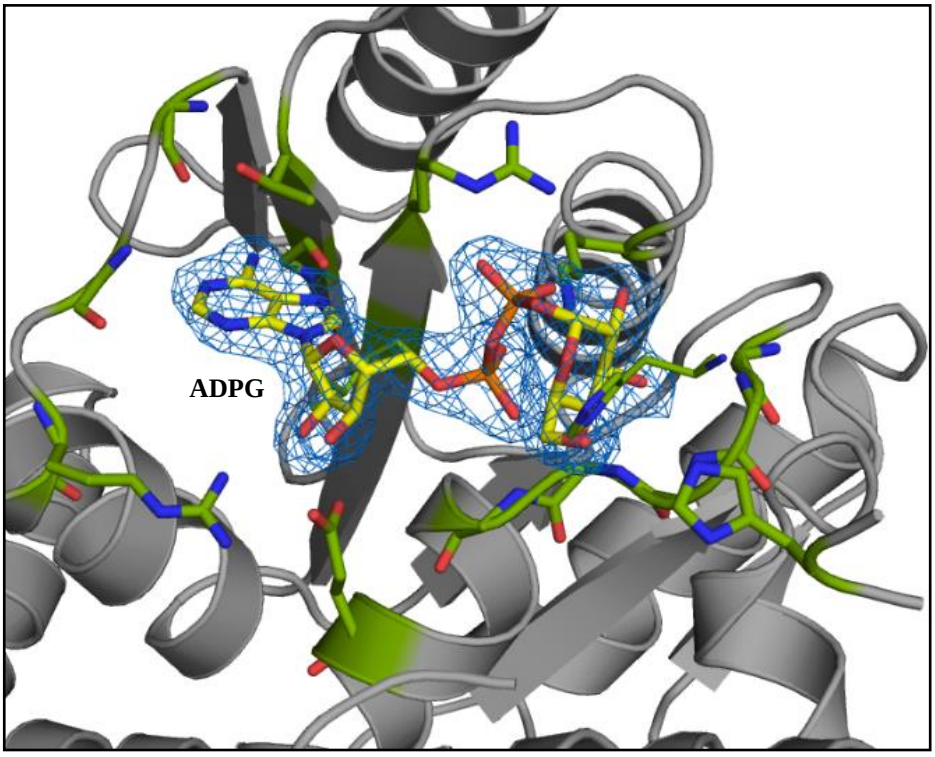
The adenine moiety of ADP-glucose occupies a deep, buried position of the catalytic site. The primary amine of the adenine moiety contacts with the carbonyl groups of Leu319 and Arg361, and N1 interacts with the amide group of Val363. The adenine rings also make π-interactions with side-chains of Val284, Thr321 and Leu386 **(Figure 2.19)**.

GDP-glucose is the next preferred substrate. The carbonyl-oxygen of the guanine-ring interacts with carbonyl and amide groups Val363. The N1 interacts with the amide group of carbonyl group of Val363, and the N7 bonds with carbonyl group of Thr321. It is probably not possible for the guanine moiety to occupy the same position as the adenine, due to steric clash of the primary amine group of the guanine-ring with the carbonyl group of Val363 **(Figure 2.20)**.

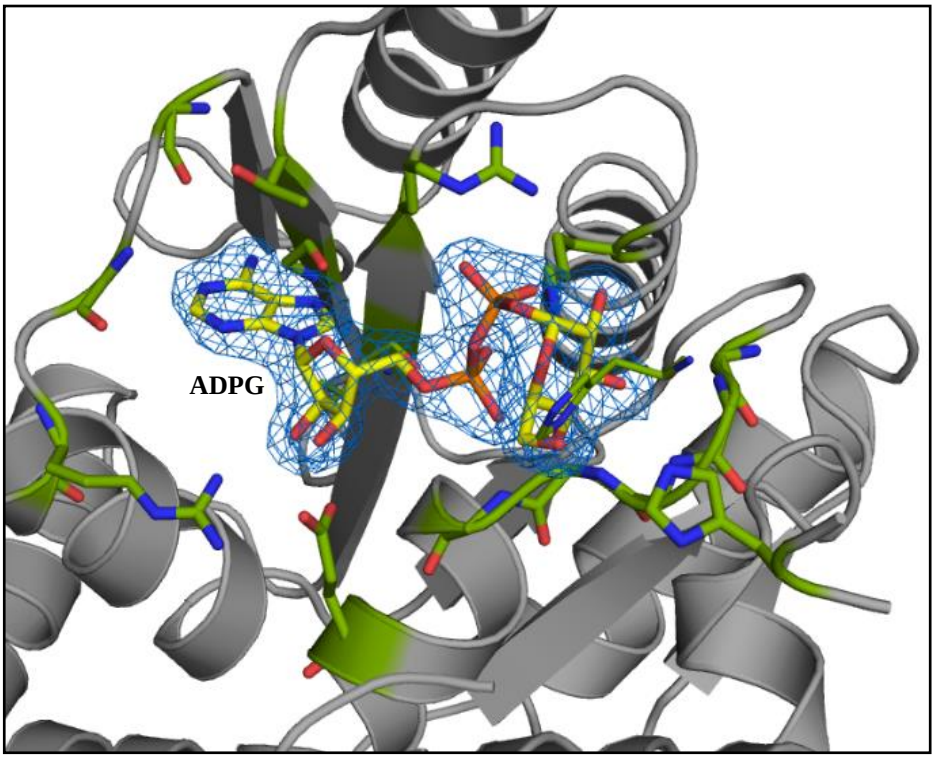
It was not possible to observe electron density of UDP-glucose in the structure due to low resolution of the crystals, and various attempts to obtain the UDP-glucose bound structure were fruitless.

The interactions ADP made with the protein were essentially similar to those made by the adenine ring, ribose sugar and phosphate groups of its activated glucose-donor **(Figure 2.21).** In the binary structure of *M. thermoresistibile* OtsA with ADP, the enzyme adopts an open conformation, and the primary amine of the adenine moiety was buried deep in the catalytic pocket, similar to what has been observed with binary structure of the enzyme with activated nucleotide donor, ADP-glucose.

A. **Protomer-1**



Protomer-2



(cont.)

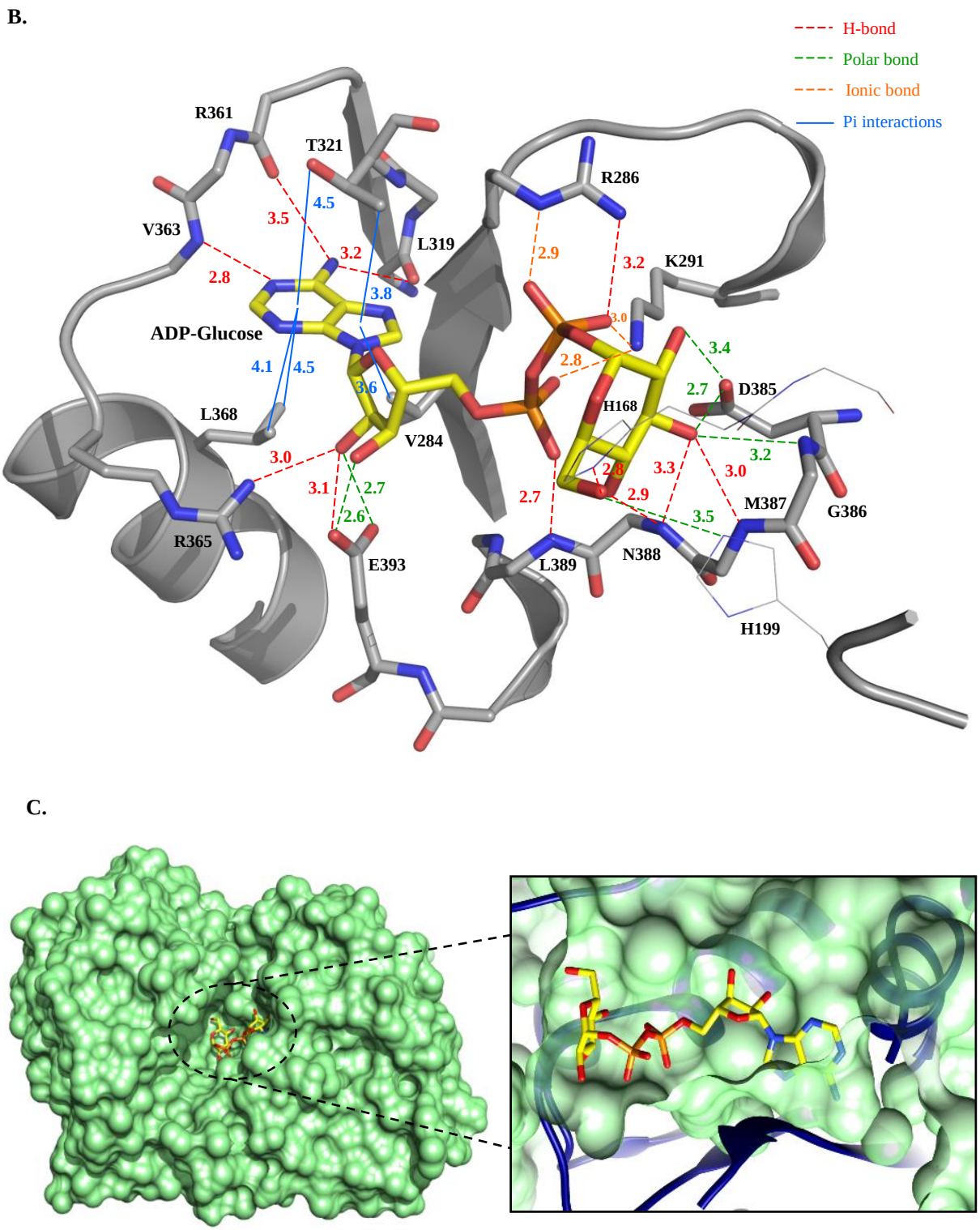


FIGURE 2.19 Structure of *M. thermoresistibile* OtsA bound with ADP-glucose.

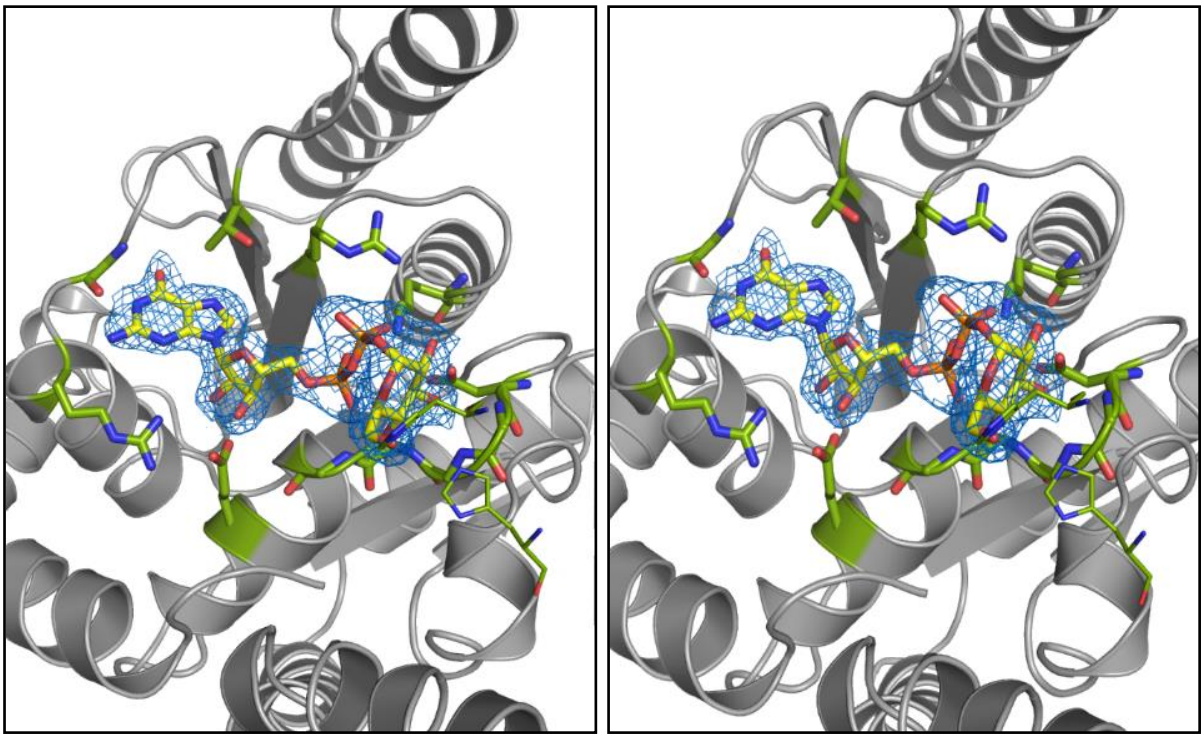
A. The observed electron density (blue mesh) for ADP-glucose (ADPG) in the catalytic site of the enzyme, shown for both the protomers in the asymmetric unit, in the same oritentation. The catalytic site residues are coloured green, but are not annotated for clarity. These are in the same orientation as in (B). The map shown is a 2Fo-Fc map, contoured at 1 σ.

B. Interaction of ADP-glucose with residues of the catalytic site of the protein. Bond lengths are shown in Å.

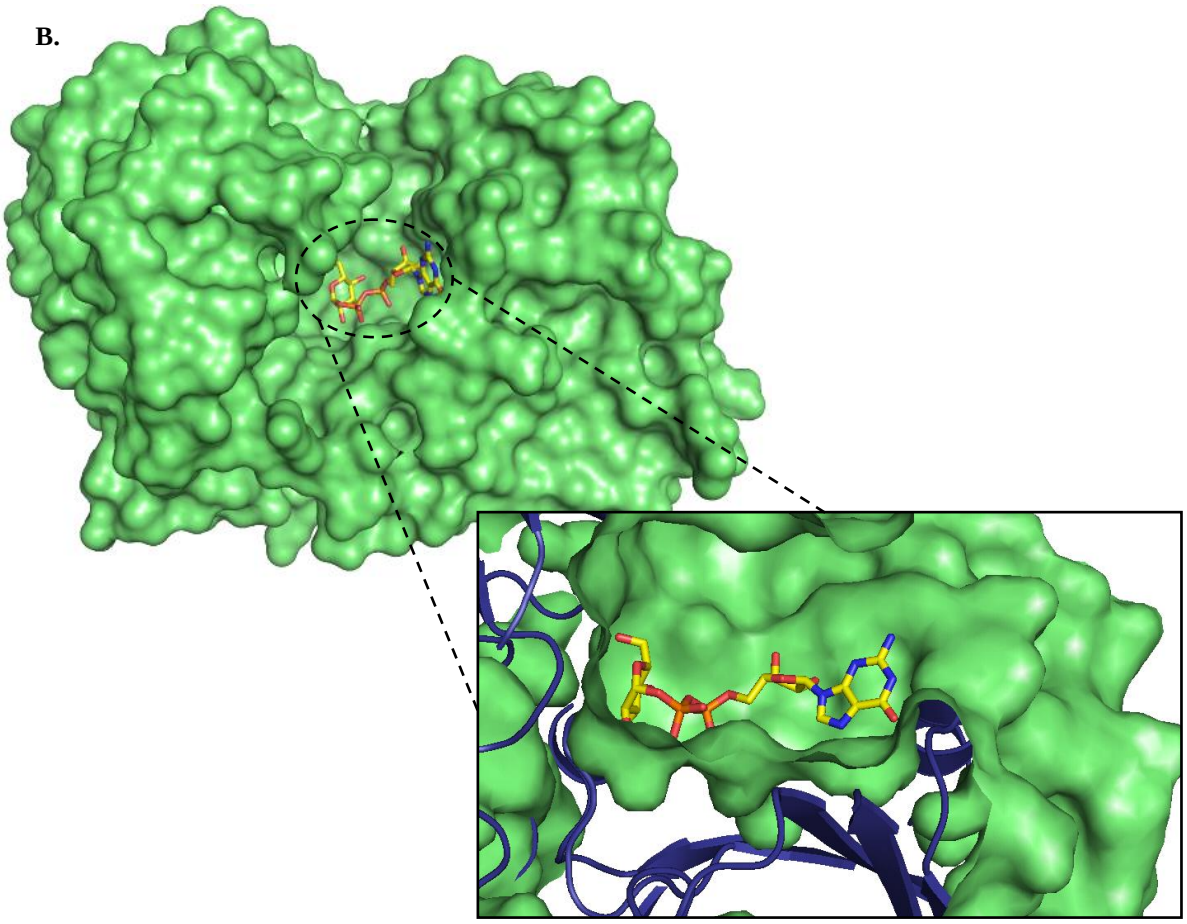
C. Surface representation of the catalytic site bound with ADP-glucose.

A.

Protomer-1 **Protomer-2**



B.



(cont.)

C.

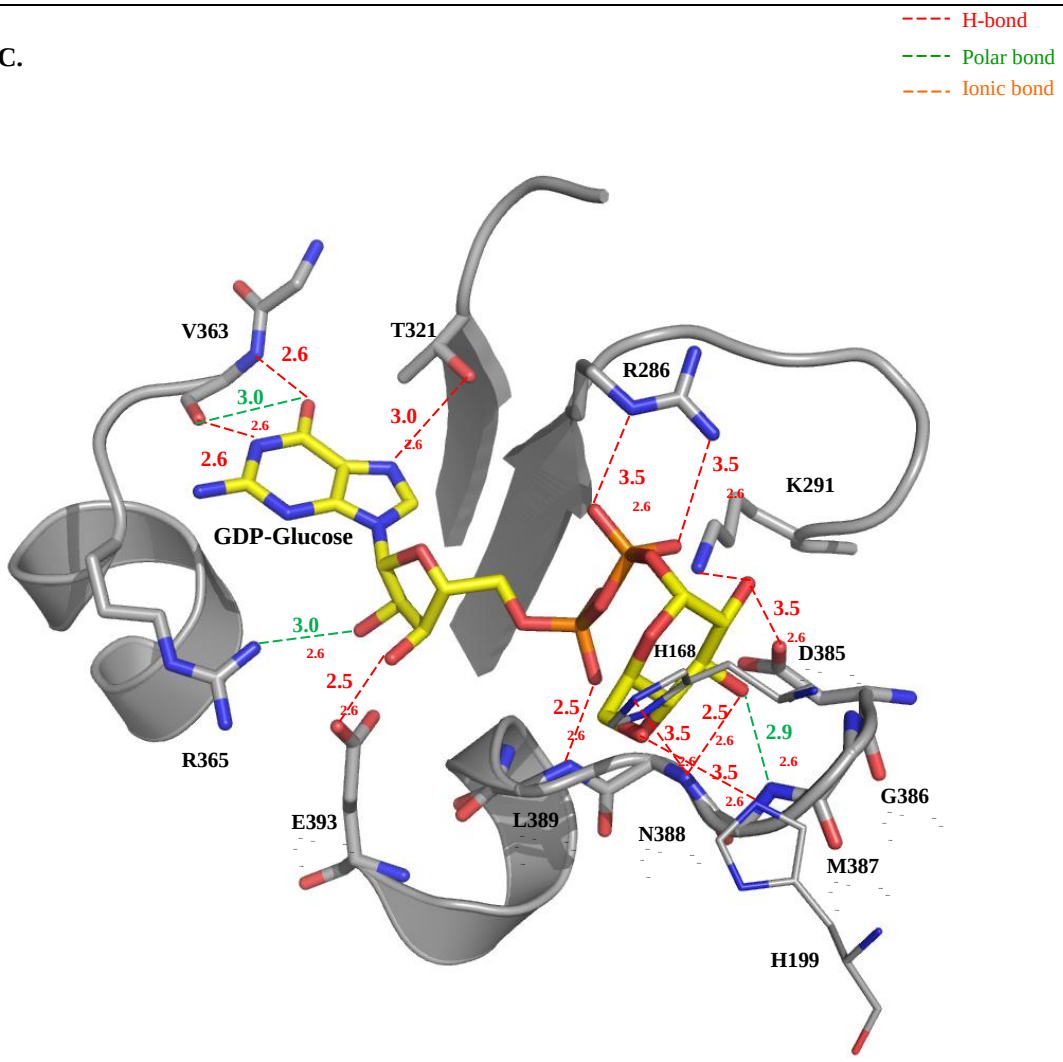


FIGURE 2.20 Structure of *M. thermoresistibile* OtsA bound with GDP-glucose.

A. The observed electron density (blue mesh) for GDP-glucose (GDPG) in the catalytic site of the enzyme, shown for both the protomers in the asymmetric unit, in the same oritentation. The catalytic site residues are coloured green, but are not annotated for clarity. These are in the same orientation as in (C). The map shown is a 2Fo-Fc map, contoured at 1 σ.

B. Surface representation of the catalytic site bound with GDP-glucose.

C. Interaction of GDP-glucose with residues of the catalytic site of the protein. Bond lengths are shown in Å.

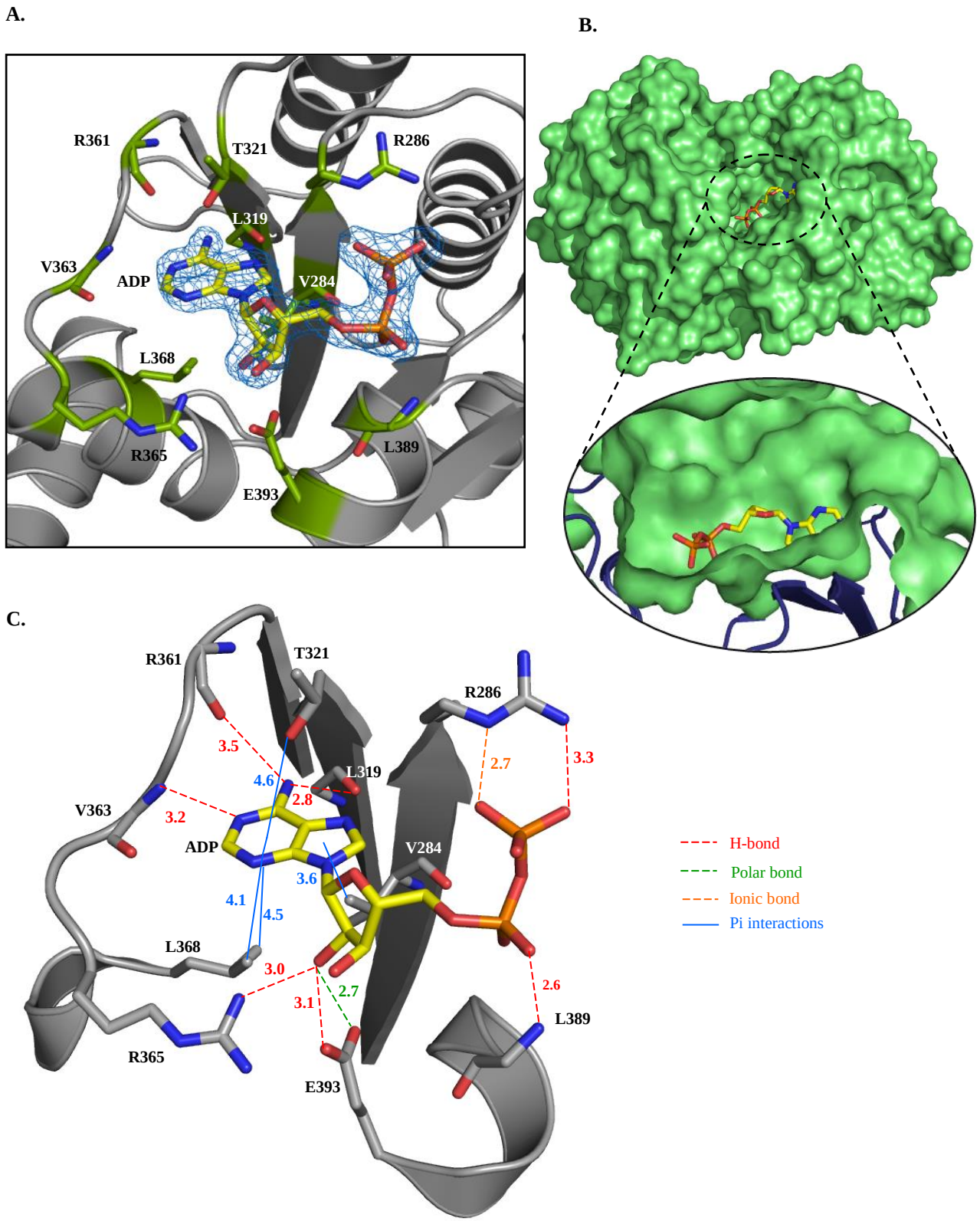


FIGURE 2.21 Structure of *M. thermoresistibile* OtsA bound with ADP.

A. The observed electron density (blue mesh) showing ADP at the catalytic site of the enzyme. The catalytic site residues are coloured green and annoted. The map shown is a 2Fo-Fc map, contoured at 1.5 σ.

B. Surface representation of the catalytic site bound with GDP-glucose.

C. Interaction of ADP with residues of the catalytic site of the protein. Bond lengths are shown in Å.

2.5.8 The Acceptor Site: Binding of Glucose-6-Phosphate and Conformational Changes Contributing to the Catalytic Mechanism of *M. thermoresistibile* OtsA

The structure of *M. thermoresistibile* OtsA with the acceptor substrate was determined by co-crystallizing the protein with glucose-6-phosphate (G6P) (glucose-acceptor substrate) and ADP (inactive donor-substrate). There was one protomer in the asymmetric unit. The data collection and refinement statistics for co-crystals are detailed in **Table 2.4**.

	ADP and G6P
Data collection*	
Beamline	DLS, I03
Space group	$P6_222$
Cell parameters:	
a/ b/ c (Å)	105.16/105.16/159.83
α/ β/ ɣ (˚)	90/90/120
Resolution range (Å)	79.92-2.14 (2.20-2.14)
No. of reflections:	
Observed	488670 (17177)
Unique	29296 (2005)
Completeness (%)	100 (99.6)
R_{meas}	0.087 (1.238)
I/ σ (I)	23.5 (2.2)
Refinement	
Refinement program	PHENIX
Resolution	79.92-2.14
No. of reflections	28793
R_{work}/R_{free}	16.1/18.36
R.M.S. Deviations:	
Bonds [Å]	0.006
Angles [˚]	1.034
Ramachandran Plot:	
Favored [%]	98.8
Outliers [%]	0

Table 2.4 Data collection and refinement parameters of ADP-G6P bound ternary *M. thermoresistibile* OtsA structure.

*Parameters in the parenthesis are for highest resolution shell

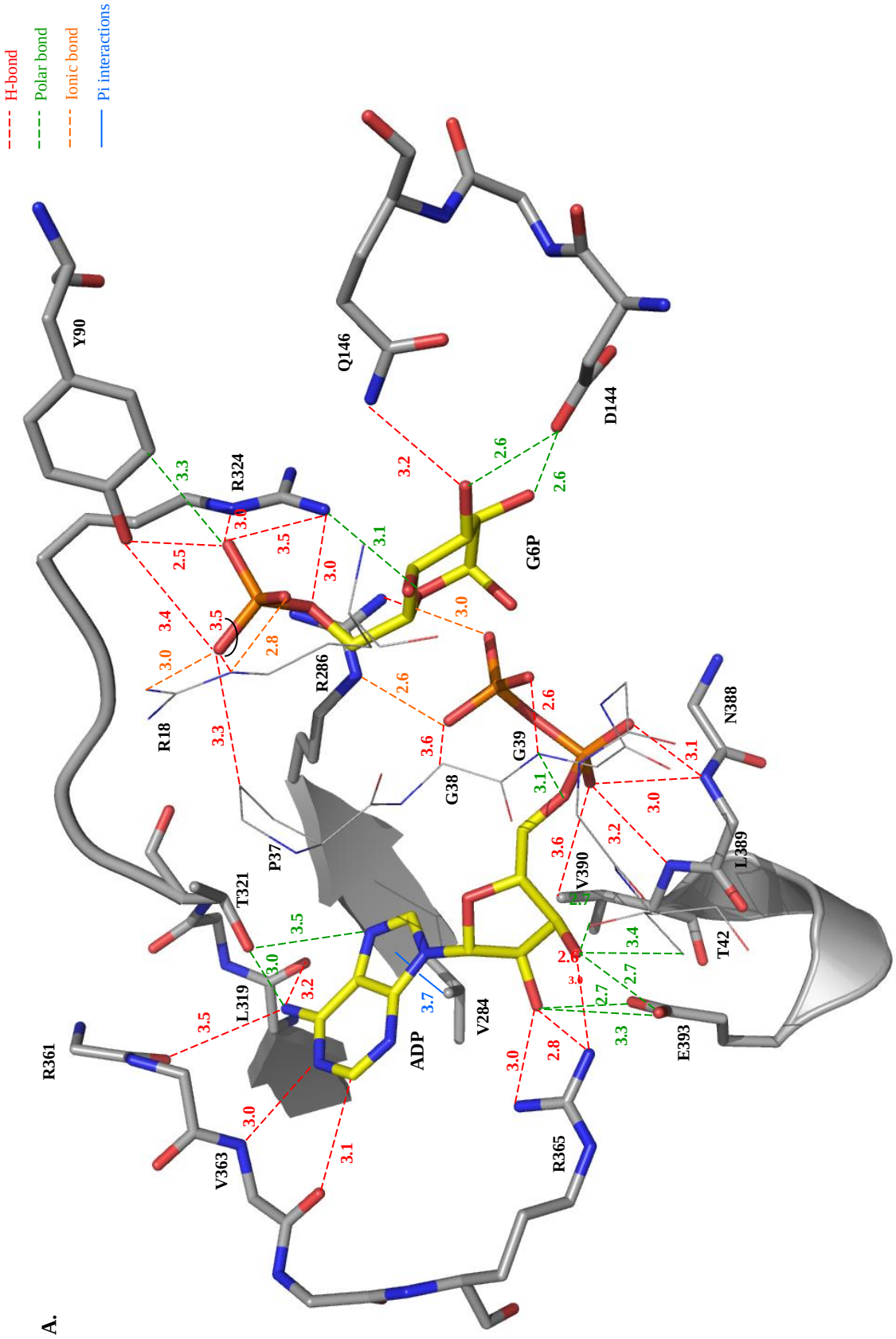
DLS: Diamond Light Source, Oxford, UK

The structure of *M. thermoresistibile* OtsA bound with ADP and G6P adopts a closed conformation. The interaction of the acceptor substrate, glucose-6-phosphate (G6P), is predominantly with the N-terminus domain through the residues Arg18, Tyr90, Asp144 and Gln146. The interaction with the C-terminus domain is through Arg324. The phosphate group bonds with Pro37, the amide groups of Arg18 and Arg324, and the carbonyl group of Tyr90. The hydroxyls of the glucose-ring coordinate with the carbonyl group of Asp144 and the amide group of Gln146. The glucose-donor substrate also establishes new contacts when the protein adopts a closed conformation, apart from the interactions it makes with the protein in the open conformation. These are interactions of the distal phosphate group with Arg18, Gly38 and Gly39, hydroxyl-group of the ribose-ring with the side chain of Thr42, and of the proximal phosphate group with Val390 **(Figure 2.22 A)**.

The ternary structure adopts the closed conformation, which is markedly different from the apo- or donor-substrate bound structures in the open-conformation **(Figure 2.23)**. The binding of G6P to the protein, along with the activated donor substrate or its nucleotide, caused the closure of the catalytic site, due to movement of α1 helix that contains the highly conserved Gly-Gly-Lue motif. Domain closure upon the binding of the acceptor substrate has also been observed for *E. coli* OtsA complexed with glucose-6-phosphate and UDP-glucose (Gibson *et al.*, 2002a). Binding of ADP alone is not able to mediate domain closure, and the protein remains in the open conformation, as observed for its activated donor, ADP-glucose (**Figures 2.19 and 2.21)**.

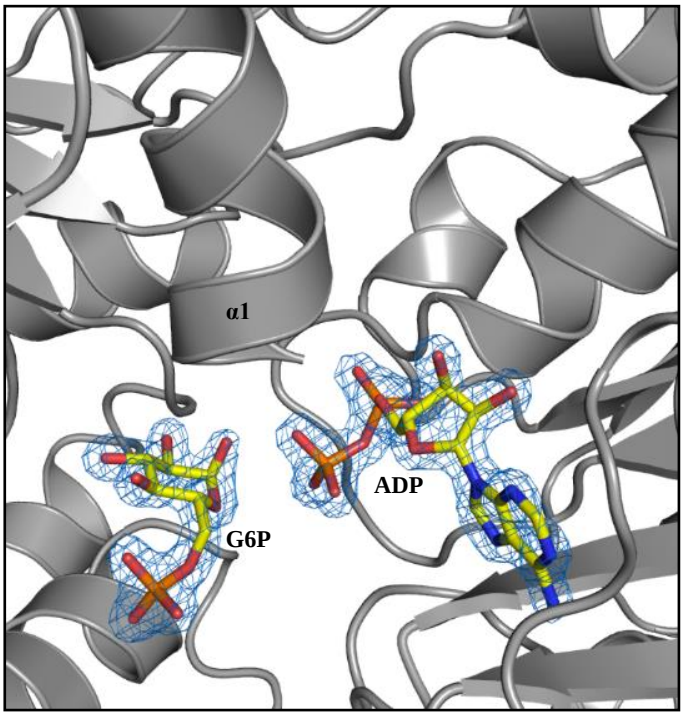
Moreover, in the OtsA:ADP:G6P ternary structure, the catalytically important loop between Arg18 and Gly39 forms a more organised β-hairpin motif that forms a part of 11-strand β-sheet at the N-terminal domain, as against the 9-strand β-sheet at the N-terminal domain in the apo-structure (**Figure 2.23)**.

The residue Gly39 is completely conserved and is most probably involved in the catalytic activity (Gibson *et al.*, 2002a). It lies on the edge of the α1 helix, and forms an interaction with the glucose-bound-phosphate of the donor substrate when the catalytic site closes. Thereby, a conformational change in the protein structure is essential for its catalytic activity.


H-bond
Polar bond
Ionic bond
Pi interactions
A.
Y90
Q146
D144
R324
G6P
R18
R286
G38
G39
N388
P37
T321
V390
L389
T42
L319
V284
ADP
E393
R361
V363
R365

(cont.)

B.



C.

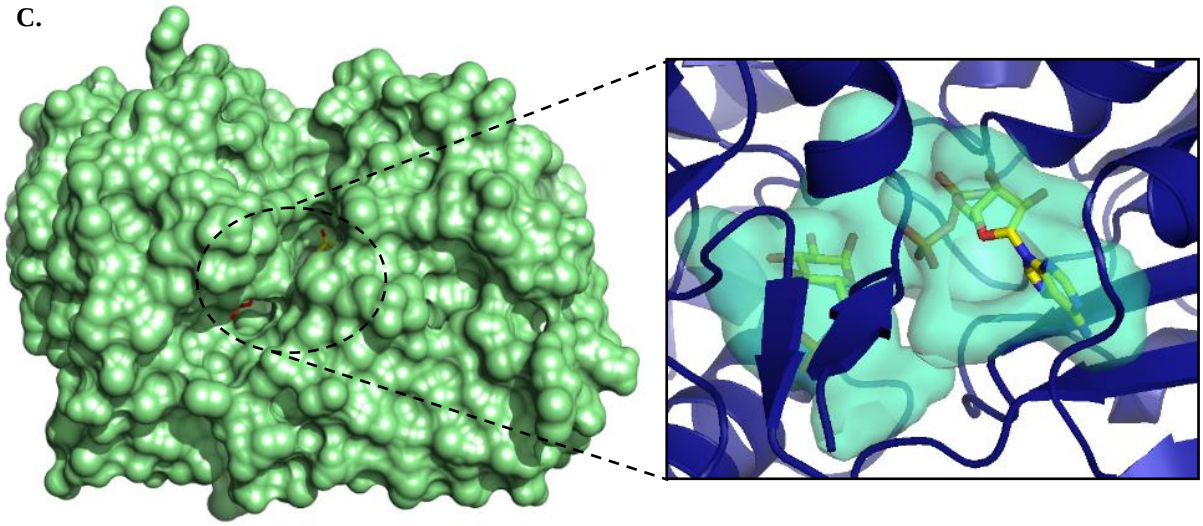


FIGURE 2.22 Ternary structure of *M. thermoresistibile* OtsA bound with ADP and glucose-6-phosphate.

A. Interactions of ADP and G6P with residues of the catalytic site of the protein. Bond lengths are shown in Å.

B. The observed electron density (blue mesh) for ADP and G6P at the catalytic centre of the enzyme. The β-hairpin motif and the interacting residues are not shown for clarity. The map shown is a 2Fo-Fc map, contoured at 1.5 σ.

C. Surface representation of the catalytic site bound with ADP and G6P. It shows the catalytic pocket in the closed conformation and the deep burial of the adenine moiety in the catalytic pocket.

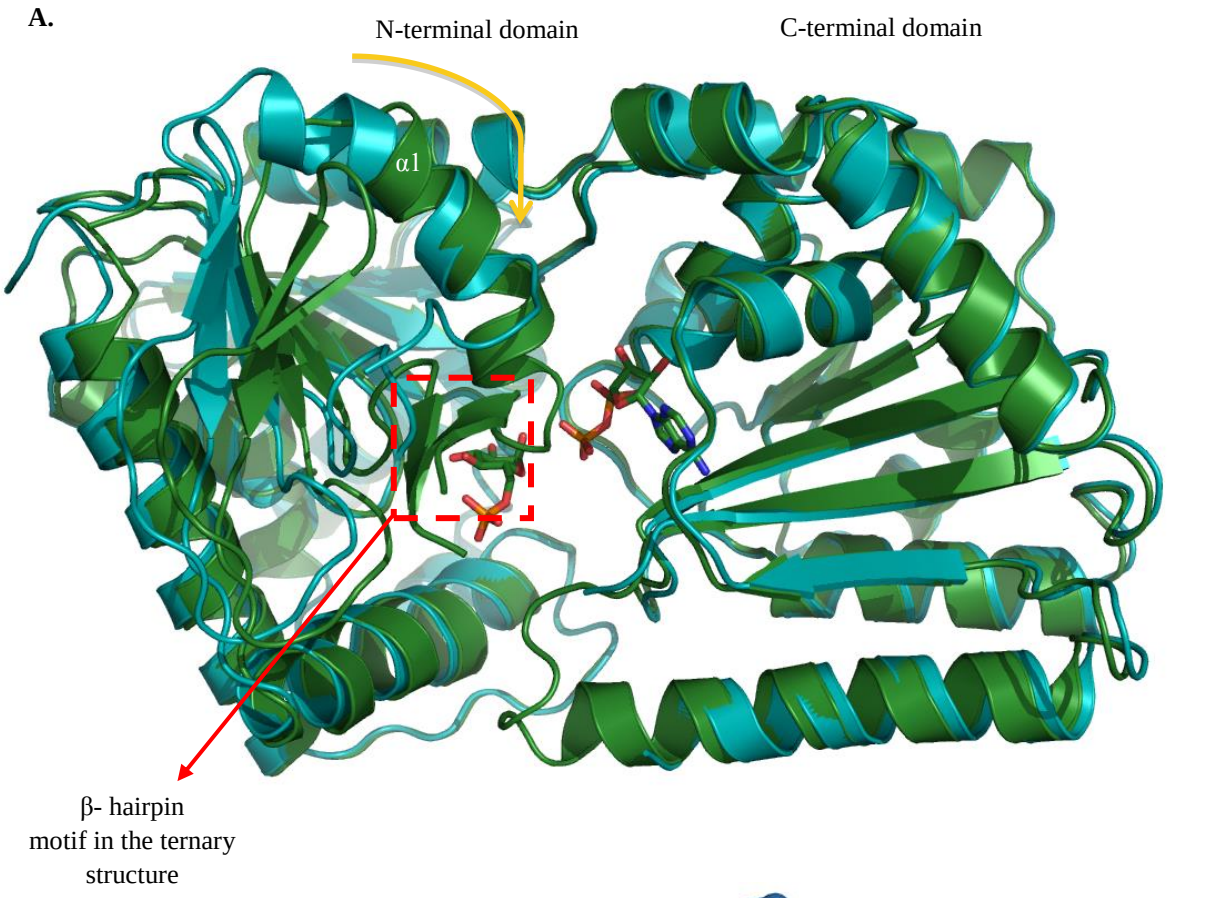


FIGURE 2.23 Conformational states of *M. thermoresistibile* OtsA.

A. Superposition of the structure of apo-protein in open conformation (teal) and OtsA-ADP-G6P ternary structure in closed conformation (green). The yellow arrow represents the direction of movement of α1 helix during the conformational change. The β-hairpin motif is marked by a dashed box.

B. Surface representation of the protein structure in the open conformation (top) and in the closed conformation (bottom). (Active site is highlighted in yellow.)

2.5.9 *M. thermoresistibile* OtsA Structures with the Products: Trehalose-6-Phosphate and Trehalose

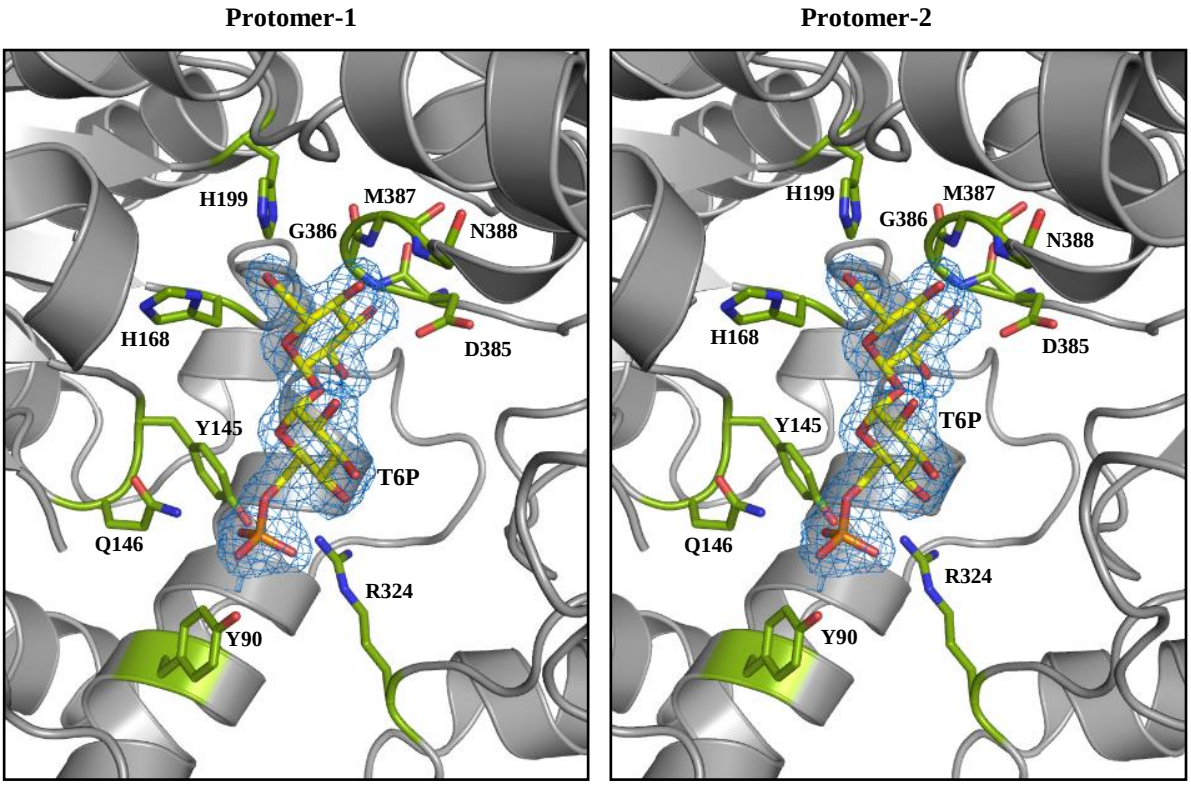
Ligand-bound structures of *M. thermoresistibile* OtsA were determined for the feedback inhibitors, trehalose-6-phosphate and trehalose, by soaking the apo-crystals with these ligands. The data collection and refinement statistics for ligand-bound crystals are detailed in **Table 2.5**. For clarity, only one protomer of the protein-ligand complex is shown in the figures. Trehalose-6-phosphate (T6P) and trehalose interact with both the N-terminal and C-terminal domains.

The phosphate group of T6P interacts with Tyr90, Tyr145, Gln146 and Arg324, similar to the interactions made by the phosphate group of G6P. The distal glucose-ring of T6P makes connection with the conserved residues His168, His199, Asp385, Gly386, Met387 and Asn388 **(Figure 2.24)**. Trehalose occupies the same position as the T6P. The distal glucose ring of trehalose makes the same contacts that the glucose-ring of T6P makes. The proximal glucose-ring of trehalose makes contact with the side-chain of Arg286 **(Figure 2.25)**.

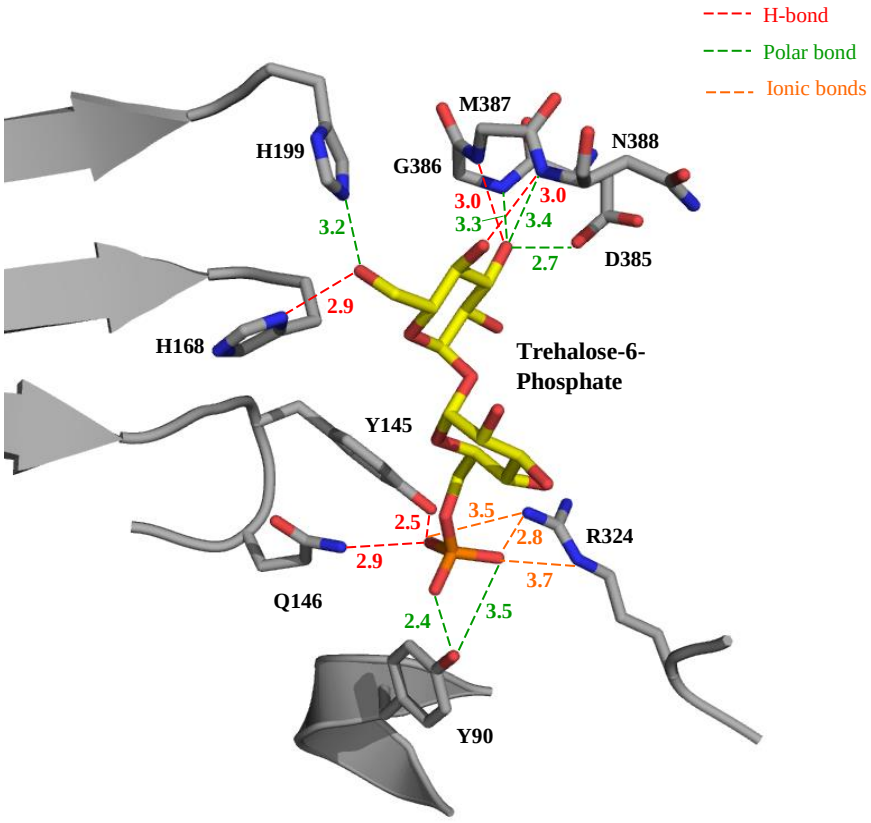
	Trehalose-6-Phosphate	**Trehalose**
Data collection*		
Beamline	DLS, I04	DLS, I04
Space group	$F4_132$	$F4_132$
Cell parameters:		
a/ b/ c (Å)	338.05/ 338.05/ 338.05	337.76/ 337.76/ 337.76
α/ β/ ɣ (˚)	90/ 90/ 90	90/ 90/ 90
Resolution range (Å)	37.8 – 2.09 (2.14 – 2.09)	57.09 – 1.97 (2.02 - 1.97)
No. of reflections:		
Observed	2462332 (124512)	2688985 (91722)
Unique	97253 (7096)	115412 (8307)
Completeness (%)	100 (100)	99.9 (99.2)
R_{meas}	0.073 (0.707)	0.064 (0.692)
I/ σ (I)	34.2 (4.6)	37.2 (3.6)
Refinement		
Refinement program	REFMAC 5	REFMAC 5
Resolution		
No. of reflections	96980	115378
R_{work}/R_{free}	16.42/17.96	18.22/20.37
R.M.S. Deviations:		
Bonds [Å]	0.010	0.008
Angles [˚]	1.035	1.041
Ramachandran Plot:		
Favored [%]	98.6	97.2
Outliers [%]	0	0

Table 2.5 Data collection and refinement parameters of trehalose-6-phosphate and trehalose bound *M. thermoresistibile* OtsA structures.
*Parameters in the parenthesis are for highest resolution shell
DLS: Diamond Light Source, Oxford, UK

A.


Protomer-1
Protomer-2
H199
M387
G386
N388
H168
D385
Y145
T6P
Q146
R324
Y90

B.


H-bond
Polar bond
Ionic bonds
M387
H199
G386
N388
3.0
3.3
3.0
3.4
3.2
2.7
D385
2.9
H168
Trehalose-6-Phosphate
Y145
3.5
2.5
2.8
R324
2.9
3.7
Q146
2.4
3.5
Y90

(cont.)

C.

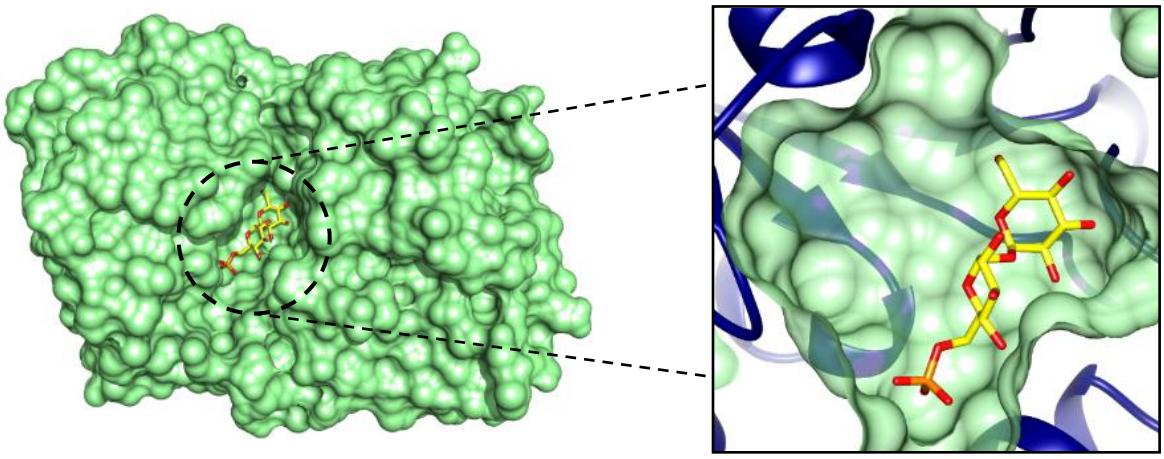


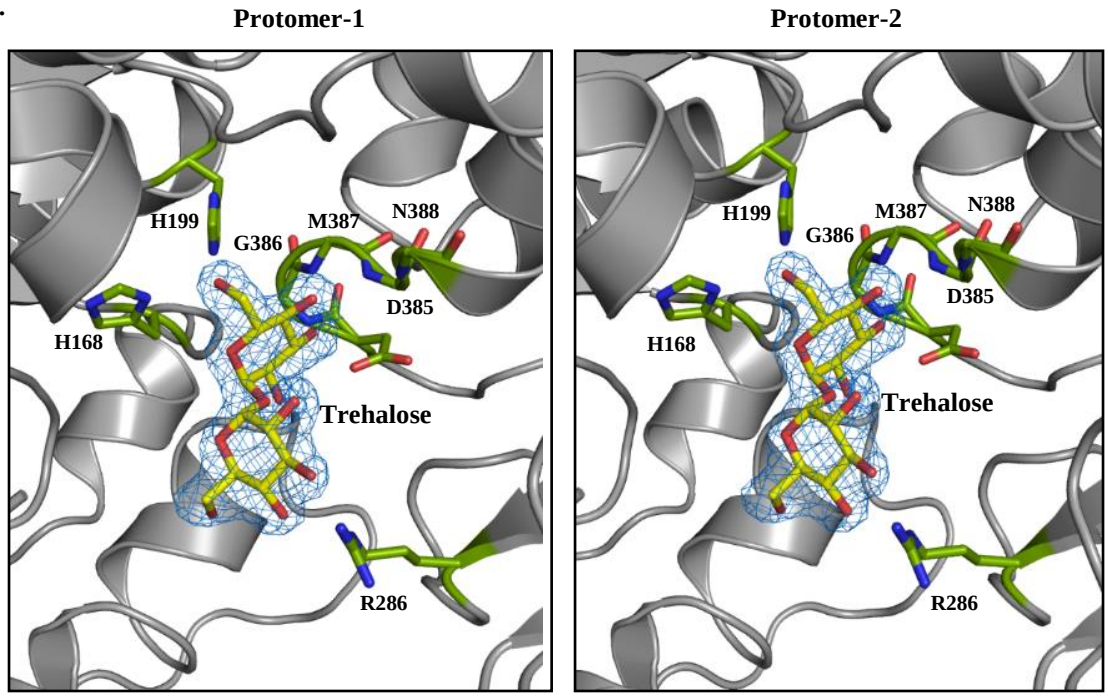
FIGURE 2.24 Structure of *M. thermoresistibile* OtsA bound with trehalose-6-phosphate.

A. The observed electron density (blue mesh) for trehalose-6-phosphate (T6P) in the catalytic site of the enzyme, shown for both the protomers in the asymmetric unit, in the same oritentation. The catalytic site residues are coloured green and annoted. The map shown is a 2Fo-Fc map, contoured at 1 σ.

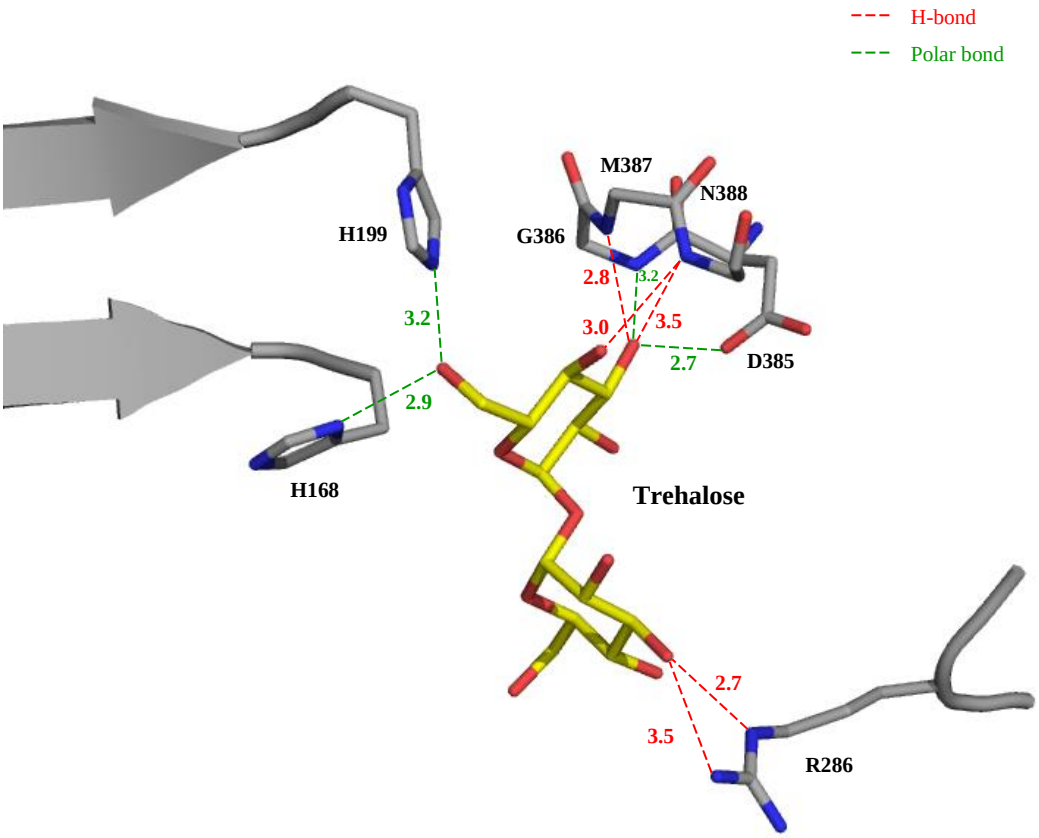
B. Interactions of trehalose-6-phosphate with residues of the catalytic site of the protein. Bond lengths are shown in Å.

C. Surface representation of the catalytic site bound with trehalose-6-phosphate.

A.


Protomer-1
Protomer-2
H199
M387
N388
G386
D385
H168
Trehalose
R286

B.


H-bond
Polar bond
M387
N388
H199
G386
2.8
3.2
3.2
3.0
3.5
2.7
D385
2.9
H168
Trehalose
2.7
3.5
R286

(cont.)

C.

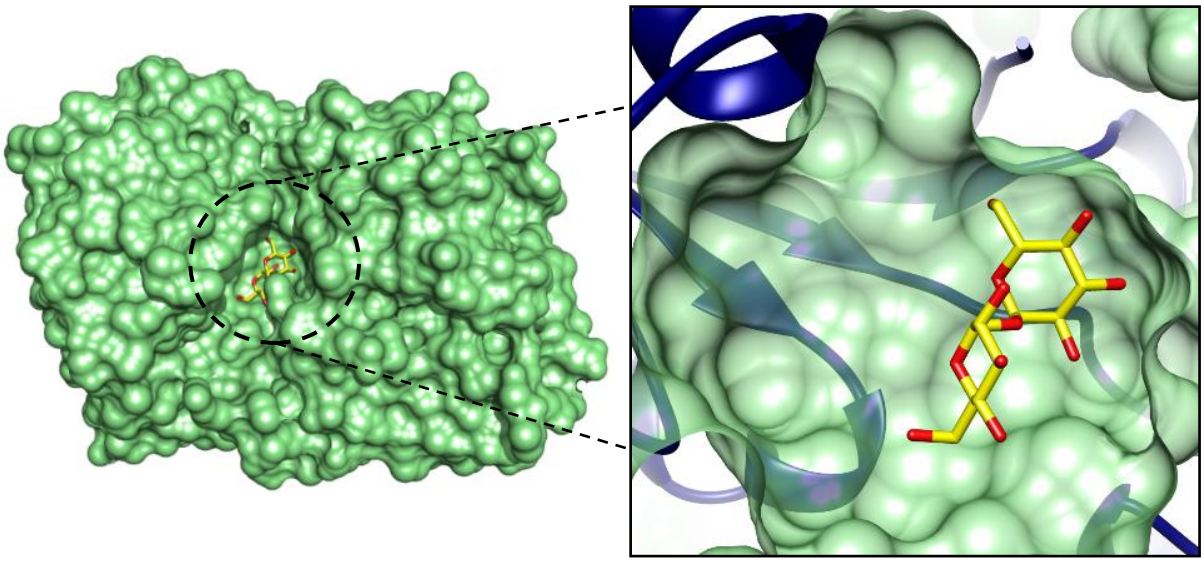


FIGURE 2.25 Structure of *M. thermoresistibile* OtsA bound with trehalose.

A. The observed electron density (blue mesh) for trehalose in the catalytic site of the enzyme, shown for both the protomers in the asymmetric unit, in the same oritentation. The catalytic site residues are coloured green and annoted. The map shown is a 2Fo-Fc map, contoured at 1 σ.

B. Interactions of trehalose with residues of the catalytic site of the protein. Bond lengths are shown in Å.

C. Surface representation of the catalytic site with trehalose bound.

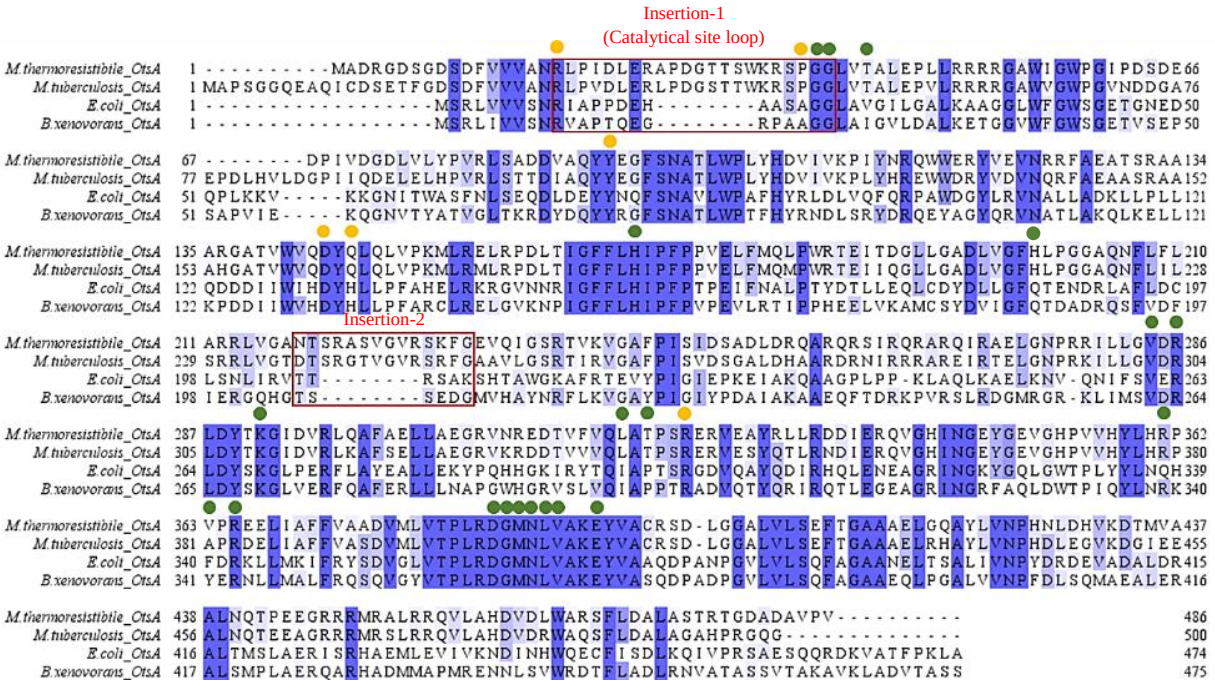
2.5.10 Comparison of Structure of Mycobacterial OtsA with Other Trehalose-6-Phosphate Synthases

A DALI search for identification of structural homologues of *M. thermoresistibile* OtsA gave the recently characterised (February, 2016) trehalose 6-phosphate synthase from *Burkholderia xenovorans* (PDB ID: 5HXA) as the top hit (Z-score =47.9), followed by *E. coli* OtsA structures bound with UDP-2-fluro-glucose (PDB ID: 1UQT, Z-score =47.5), and with UDP-glucose (PDB ID: 1UQU, Z-score= 47.4).

M. thermoresistibile OtsA shares 34% sequence identity with *E. coli* and 35% identity with *B. xenovorans*. The structural alignment of mycobacterial OtsA with these OtsA structures shows that the overall structural fold is similar, quintessential of GT-B glycosyltransferases **(Figure 2.26)**. The structural alignment between mycobacterial OtsA and *E. coli* OtsA (PDB

ID: 1UQU) resulted in RMSD of 1.55 Å over 415 residues, whereas the structural alignment between mycobacterial OtsA and *B. xenovorans* OtsA (PDB ID: 5HXA) resulted in RMSD of 1.45 Å over 418 residues. The comparative analysis shows that mycobacterial OtsA has insertions in two loops. The first insertion is of 8 residues in the catalytic site loop between Arg18 and Gly39. The second insertion is also of 8 residues in the loop between Asn218 and Glu232.

A.



(cont.)

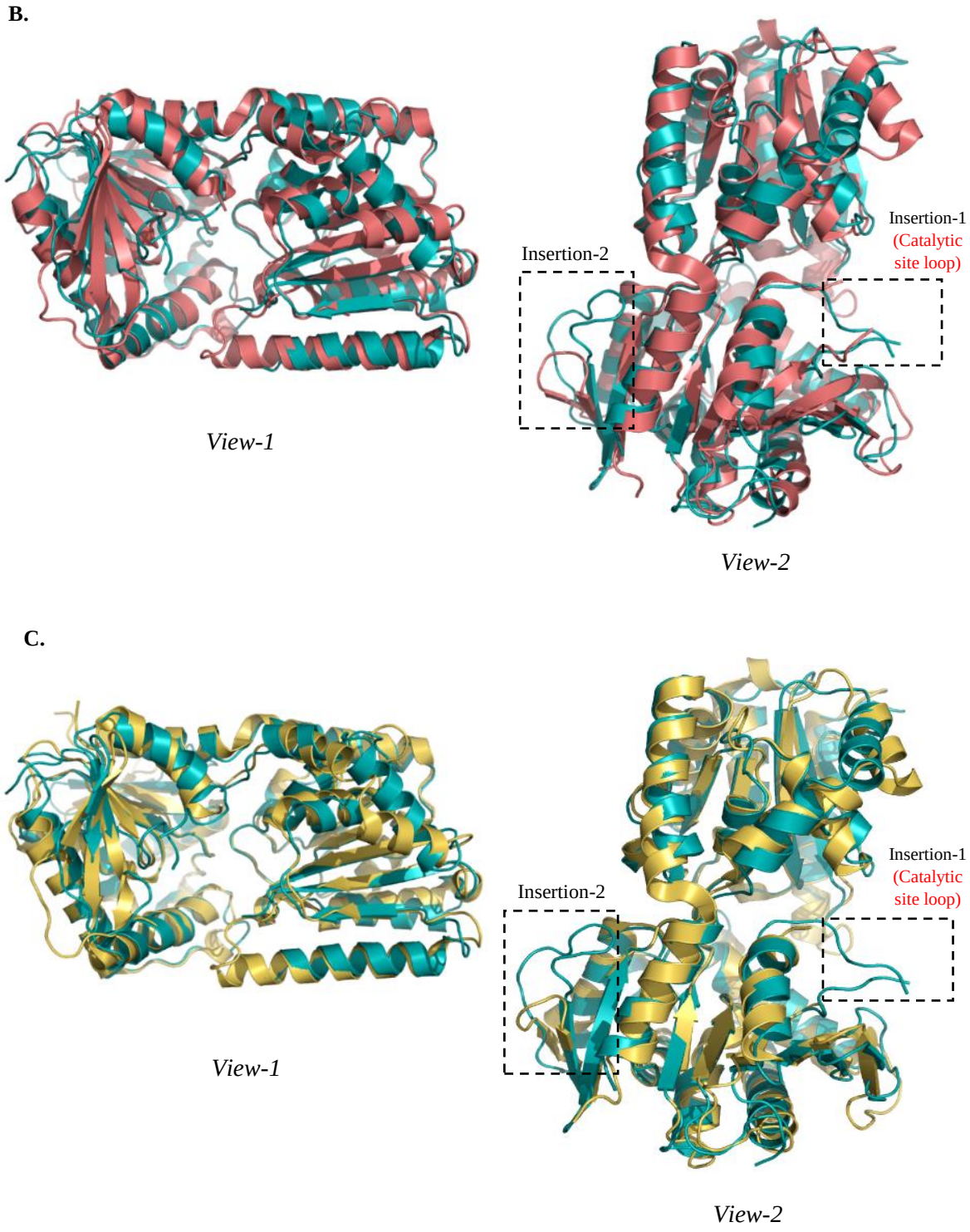


FIGURE 2.26 Comparison of structures of *M. thermoresistibile* OtsA with those of other OtsA proteins.

A. Sequence alignment of *M. thermoresistibile* OtsA with *E. coli* and *B. xenovorans* OtsA, showing the insertions in *M. thermoresistibile* OtsA. Residues of catalytic site of *M. thermoresistibile* OtsA are marked with circles (yellow: acceptor substrate site, green: donor substrate site).

B. Superposition of structures of *M. thermoresistibile* OtsA (teal) and *E. coli* OtsA (PDB ID: 1UQU) (pink) in open conformation.

C. Superposition of structures of *M. thermoresistibile* OtsA (teal) and *B. xenovorans* OtsA (PDB ID: 5HXA) (golden) in open conformation.

Comparison of the Ternary Structures in the Closed-Conformation

The ternary complexes of *M. thermoresistibile* OtsA:ADP:G6P, and *E. coli* OtsA:UDP:G6P (PDB ID: 1GZ5) adopt a closed conformation. Superposition of these structures showed that the α1 helix of *M. thermoresistibile* OtsA fully overlaps with the helix of *E. coli* OtsA, and the conformational change is restricted to the N-terminus acceptor-domain. In *M. thermoresistibile* OtsA, the catalytic-site loop having an insertion, forms an organised β-hairpin motif in the ternary structure, which is not observed in *E. coli* OtsA ternary structure **(Figure 2.27)**.

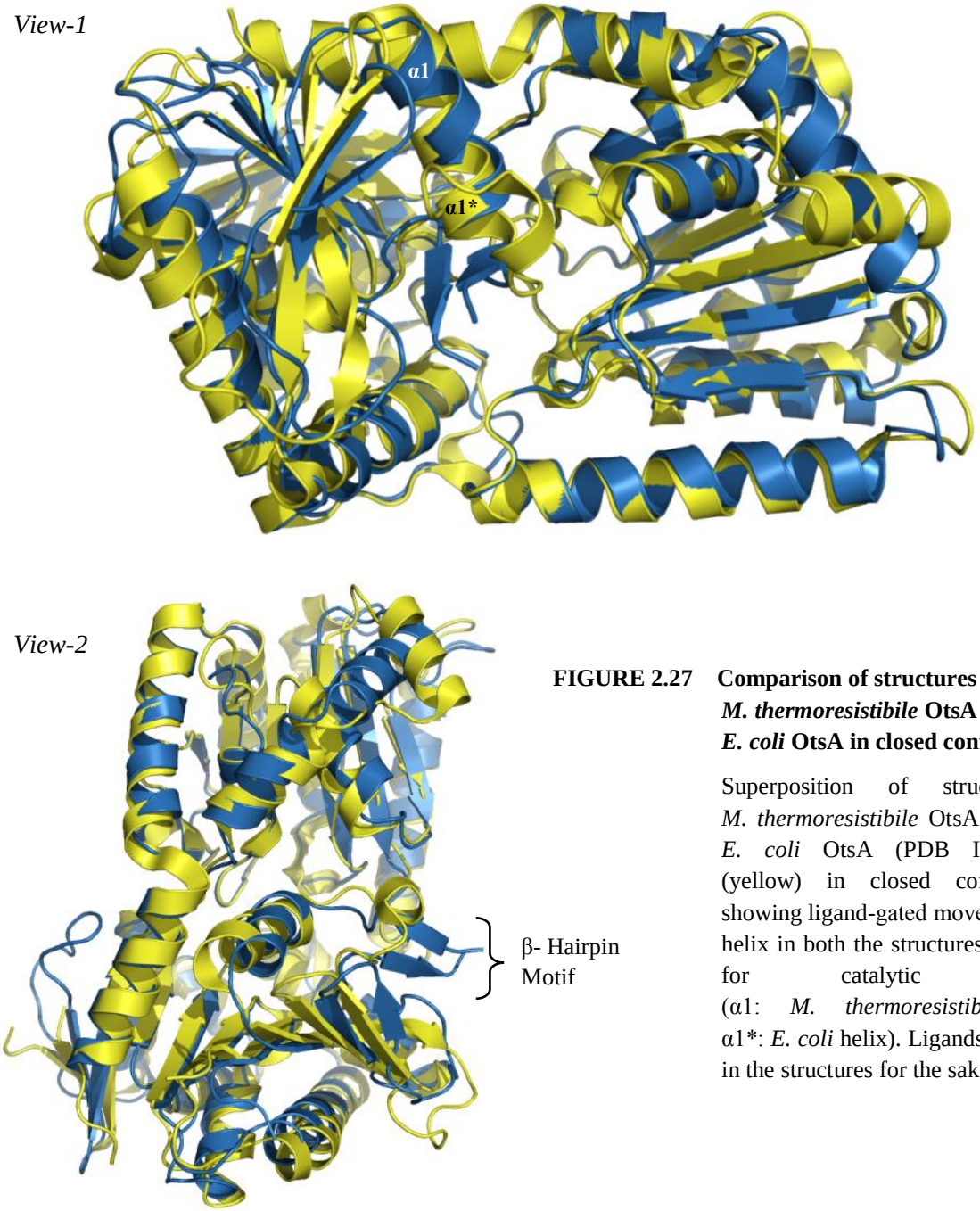


FIGURE 2.27 Comparison of structures of *M. thermoresistibile* OtsA with *E. coli* OtsA in closed conformation.

Superposition of structures of *M. thermoresistibile* OtsA (blue) and *E. coli* OtsA (PDB ID: 1GZ5) (yellow) in closed conformation, showing ligand-gated movement of α1 helix in both the structures, necessary for catalytic activity. (α1: *M. thermoresistibile* helix; α1*: *E. coli* helix). Ligands not shown in the structures for the sake of clarity.

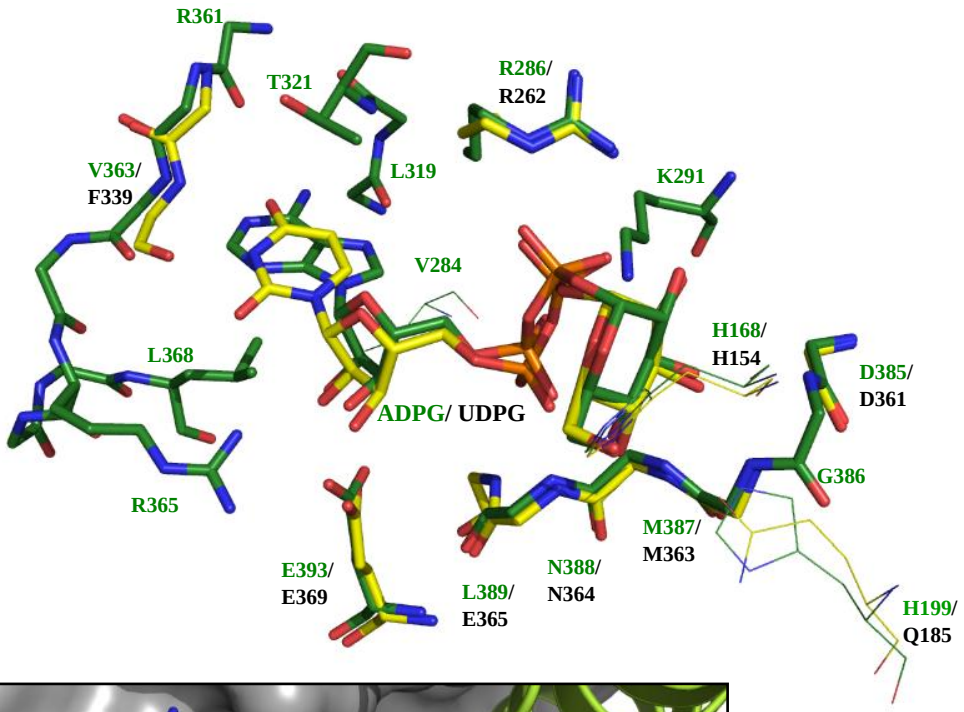
Comparison of the Nucleotide Binding Site

The donor-substrate bound structures of *M. thermoresistibile* and *E. coli* OtsA adopt an open conformation. In both the enzymes, the residues interacting with the glucose rings and phosphate groups of the NDP-glucoses are mostly conserved. The ribose, phosphate and glucose moieties of ADP-glucose or GDP-glucose in the catalytic pocket of *M. thermoresistibile* OtsA are respectively superimposable onto the ribose, phosphate and glucose moieties of UDP-glucose in the catalytic pocket of *E. coli* OtsA. However, the position occupied by different nitrogenous bases and the interactions they make with the enzyme from respective species are different **(Figure 2.28 A)**.

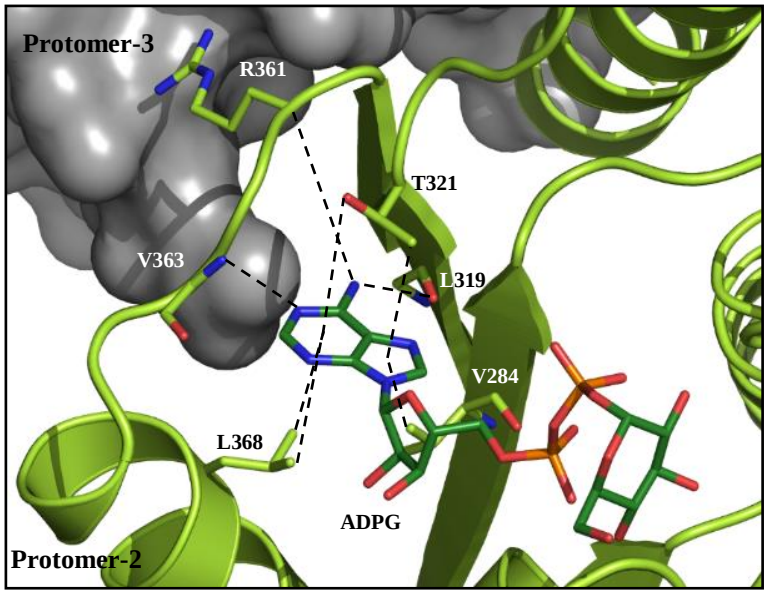
Gibson and colleagues have shown that in the binary *E. coli* OtsA:UDP-glucose structure, O4 and N3 of the uracil moiety make contact with hydrophobic Phe339, and because there are no interactions with O2, it is possible to model guanine in place of uracil (Gibson *et al*., 2002a). On the other hand, structures of *M. thermoresistibile* OtsA with the donor substrates described in this chapter show that the adenine or guanine moiety of ADP-glucose or GDP-glucose respectively, interact with multiple residues: Val 284, Leu319, Thr321, Arg361, Val363 and Leu368 **(Figure 2.28 B)**.

The comparison of the catalytic pocket of *E. coli* and *M. thermoresistibile* OtsA reveals that the divergent substrate preferences may be attributed to several amino-acid substitutions. The preference for ADP-glucose by *M. thermoresistibile* OtsA (instead of uracil by *E. coli* OtsA) is most likely due to the substitution of Ile295 and Phe339 in *E. coli* OtsA by Leu319 and Val363 respectively in *M. thermoresistibile* OtsA. In both *M. thermoresistibile* and *E. coli* OtsA, Leu319 and Ile295 respectively occupy a similar position on a loop in proximity to the adjacent protomer; and Val363 and Phe339 occupy a position on a β-strand. The catalytic site in *M. thermoresistibile* OtsA is enlarged by these substitutions that favour the bicyclic adenine moiety, especially its primary amine group, to closely-fit in the pocket **(Figure 2.28 C)** and make extensive contacts with the hydrophobic and polar residues surrounding it. This is consistent with the high activity of *M. thermoresistibile* with ADP-glucose. Within *M. thermoresistibile* OtsA, the positioning of guanine of GDP-glucose at this site would be prevented due to possible steric clash of its primary amine with the carbonyl of Val363 (explained in the section 2.5.7, figure 2.20).

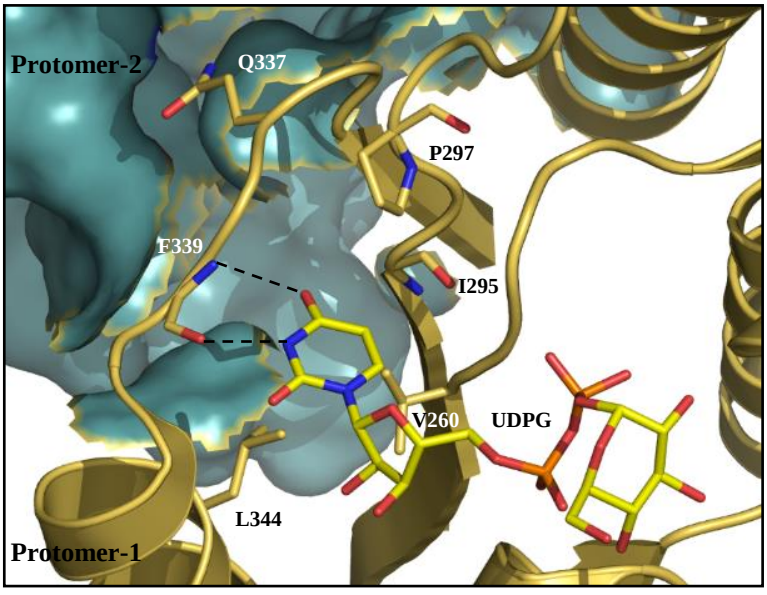
A.



B.



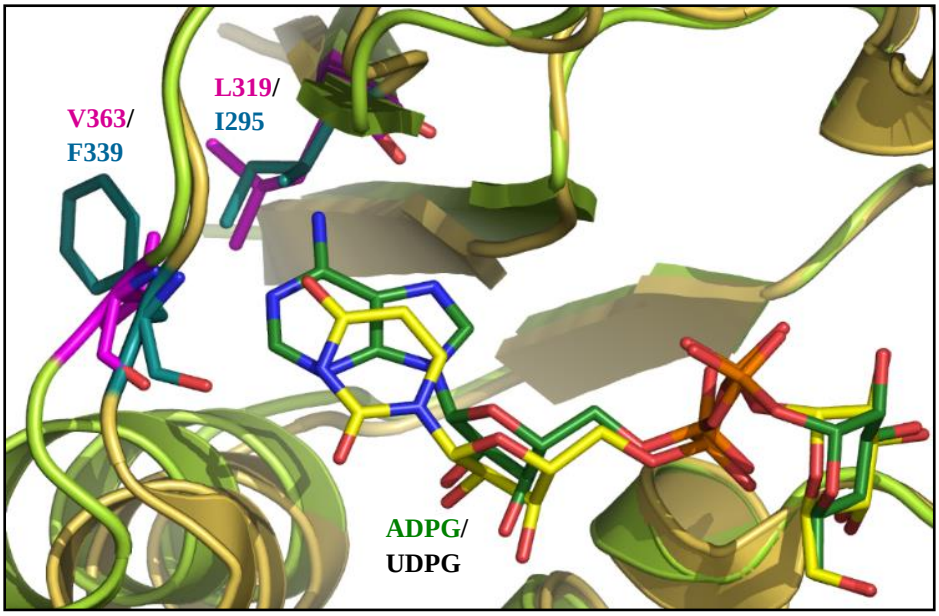
M. thermoresistibile OtsA



E. coli OtsA

(cont.)

C.



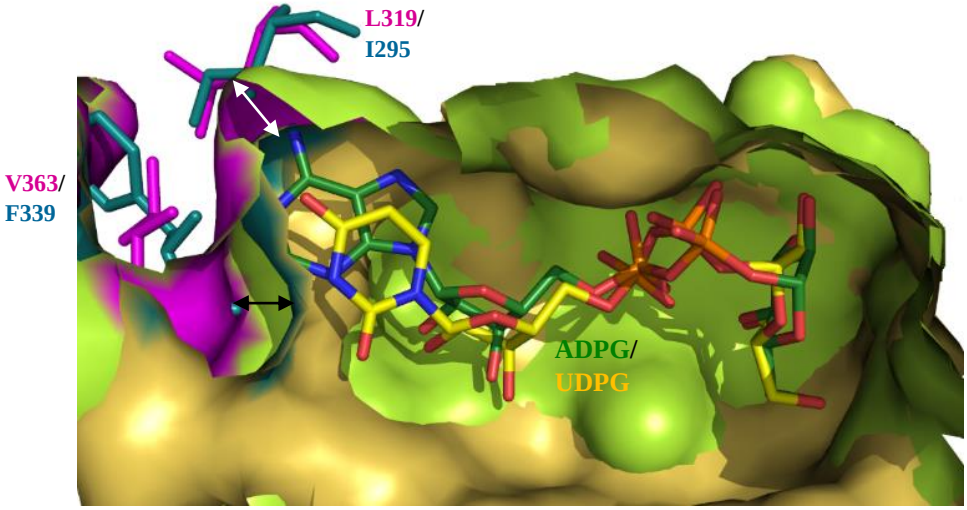


FIGURE 2.28 Comparison of the nucleotide-donor binding sites of *M. thermoresistibile* OtsA with *E. coli* OtsA.

B. Superposition of activated nucleotide donor site of *M. thermoresistibile* OtsA (green) and *E. coli* OtsA (Gibson *et al*., 2004; PDB ID: 1UQU) (yellow). Note the differences in the position of the nitrogenous bases. Only one protomer is shown for each protein. The interacting residues of *M. thermoresistibile* OtsA are labelled in green, wherase those of *E. coli* OtsA are labelled in black.

C. Orthoscopic view of the activated nucleotide donor sites of *M. thermoresistibile* OtsA and *E. coli* OtsA in the same orientation, focusing on the interactions made by the nitrogenous bases. The adjacent protomer is shown in the molecular surface representation.

D. Activated nucleotide donor sites of *M. thermoresistibile* OtsA (lime green) and *E. coli* OtsA (golden) in cartoon representation (top) and molecular surface representation (bottom), superposed to shown the differences made by the amino-acid substitutions. L319 and V363 of *M. thermoresistibile* OtsA are coloured magenta, and I295 and F339 of *E. coli* OtsA are coloured teal. Note the deepening of the catalytic pocket of *M. thermoresistibile* OtsA by the substitutions, marked by arrows.
(ADPG: ADP-glucose, UDPG: UDP-glucose)

2.5.11 Fragment-Based Drug Discovery Against *M. thermoresistibile* OtsA

2.5.11.1 Fluorescence-Based Thermal Shift Assay as a High Throughput Preliminary Screening Method for Fragment Hits

The preliminary screening of the fragment library was carried out using a fluorescence-based thermal shift assay. A fragment is considered a hit, if upon its binding, the melting temperature of the enzyme is altered. If the binding of a fragment increases the melting temperature of the enzyme, then it is known as a positive hit; and conversely, if the melting temperature is reduced, then the fragment is known as a negative hit (Scott *et al*., 2012).

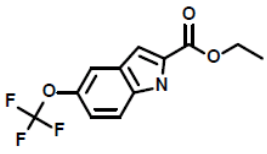
The effect of DMSO on the stability of *M. thermoresistibile* OtsA was assessed prior to the fragment screening, as the fragments were dissolved in DMSO. The average melting temperature of the enzyme was determined in the absence and presence of DMSO, and was observed to be 63.50 ± 0˚C and 61.75 ± 0.30˚C, respectively. This indicated that DMSO had a modest destabilisation effect on the enzyme stability. The binding of ADP-glucose (dissolved in DMSO) to the enzyme, had a pronounced effect on the thermal stability of enzyme, and increased its melting temperature by 19.20 ± 0.25 ˚C, whereas, ADP (dissolved in DMSO) gave a minimal positive shift of +1˚C. A total of 960 fragments were tested and two types of screening processes were carried out for identification of hits against *M. thermoresistibile* OtsA.

In type-1 screening, the fragment library was screened against the enzyme in the apo-state, i.e in the absence of any other ligand. The melting temperature of protein in the presence of DMSO (61.75 ± 0.30˚C), and that with ADP-glucose (80.88 ± 0.25˚C) were used as negative and positive controls, respectively. A fragment was considered as a hit from type-1 screen, when its binding to the enzyme caused a change in the melting point by ± 1.5°C, compared with the negative control. This cut-off value was five-times the value of the standard deviation of the measured melting point of the negative control.

The majority of hits from this screening process were negative, and had a destabilisation effect on the protein. This may possibily be due the terameric organisation of the enzyme, which gets disrupted upon binding of certain fragments, and result in negative shifts. A total of 15 fragments were identified as positive hits, giving a hit rate of 1.56%. The chemical

structures and value of ΔTm of these positive hits, along with representative thermal profiles are presented in **Figure 2.29**.

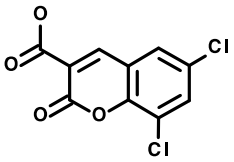
A.



F1.

Ethyl 5-(trifluoromethoxy) indole-2-carboxylate

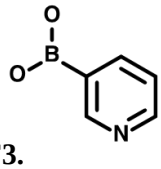
ΔTm= +16˚C



F2.

6,8-dichloro-2-oxo-2H-chromene-3-carboxylic acid

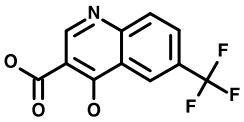
ΔTm= +15.5˚C



F3.

3-Pyridinylboronic acid

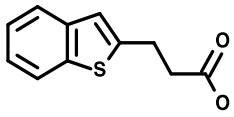
ΔTm= +15.38˚C



F4.

4-Hydroxy-6-(trifluoromethyl)-3-quinolinecarboxylic acid

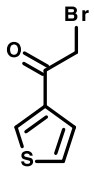
ΔTm= +15.13˚C



F5.

Benzo[b]thiophene-2-propionic acid

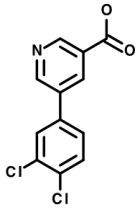
ΔTm= +11.13˚C



F6.

2-bromo-1-(3-thienyl)-1-ethanone

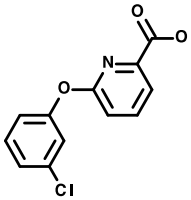
ΔTm= +11˚C



F7.

5-(3,4-Dichlorophenyl) nicotinic acid

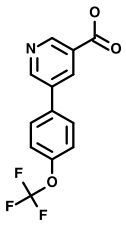
ΔTm= +9.5˚C



F8.

6-(3-chlorophenoxy) pyridine-2-carboxylic acid

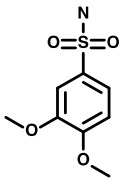
ΔTm= +7.5˚C



F9.

5-(4-Trifluoromethoxyphenyl)-nicotinic acid

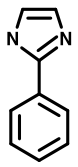
ΔTm= +6.63˚C



F10.

3,4-Dimethoxybenzene sulfonamide

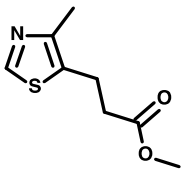
ΔTm= +1.75˚C



F11.

2-Phenylimidazole

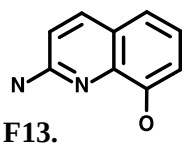
ΔTm= +1.75˚C



F12.

2-(4-Methyl-5-thiazolyl) ethyl acetate

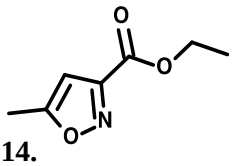
ΔTm= +1.75˚C



F13.

2-Amino-8-quinolinol

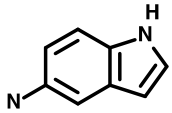
ΔTm= +1.75˚C



F14.

Ethyl-5-methylisoxazole-3-carboxylate

ΔTm= +1.75˚C



F15.

5-Aminoindole

ΔTm= +1.38˚C

(cont.)

B.

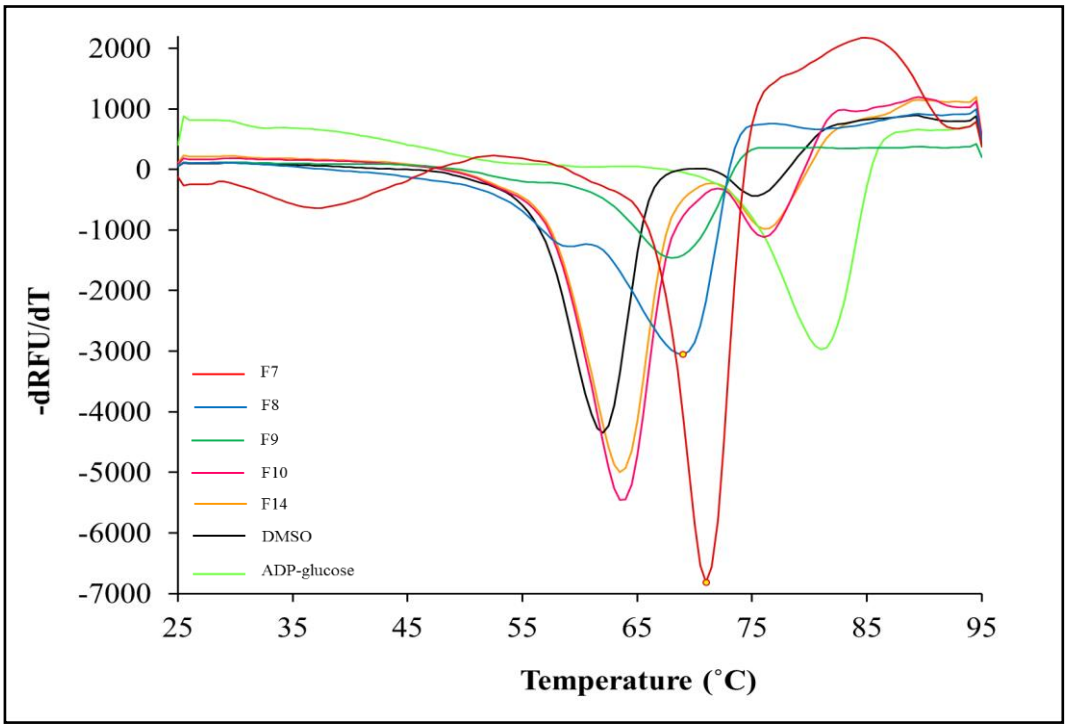
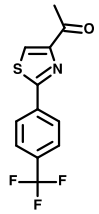


FIGURE 2.29 Fragment hits from type-1 (in the absence of ADP) screening of the fragment library using thermal shift assay.

A. Chemical structures and corresponding ΔTm values of the fragment hits.

B. Representative thermal shift profiles of some fragment hits.

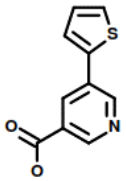
The structural analyses of *M. thermoresistibile* OtsA demonstrated that the enzyme exhibits different conformational states. This structural knowledge was exploited, and formed the rationale for screening the fragment library in the presence of ADP, to identify the fragments that were able to modulate the conformational change. In this type-2 screening process, the same principles were applied and analysis was done as for the type-1 screening. The melting point of the enzyme in presence of ADP (62.50 ± 0˚C) was taken as the negative control, and that of protein with ADP-glucose (81.50 ± 0.70˚C) served as the positive reference. The cut-off value for classification of fragments as a hit was ±1˚C. The majority of hits from the type-2 screening were negative, as observed for type-1 screen. From this screening, 7 fragments were classified as hits (hit rate=0.73%), and the maximum value of positive ΔTm observed was +15˚C. The chemical structures and value of ΔTm of these 7 positive hits are given in **Figure 2.30**.



F1.

1-{2-[4-(trifluoromethyl)phenyl]-1,3-thiazol-4-yl}ethan-1-one

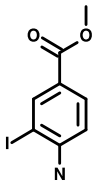
ΔTm= +15˚C



F2.

5-(2-thienyl) nicotinic acid

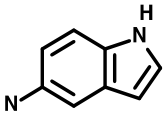
ΔTm= +14C



F3.

Methyl 4-amino-3-iodobenzoate

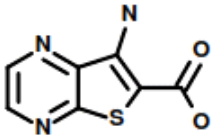
ΔTm= +3.88˚C



F4.

5-Aminoindole

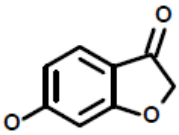
ΔTm= +2.5˚C



F5.

7-aminothieno[2,3-b] pyrazine-6-carboxylic acid

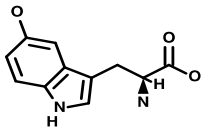
ΔTm= +2˚C



F6.

6-hydroxy-2,3-dihydrobenzo[b]furan-3-one

ΔTm= +1.5˚C



F7.

5-Hydroxy-L-tryptophan

ΔTm= +1.25˚C

FIGURE 2.30 Fragment hits from type-2 (in the presence of ADP) screening of the fragment library using thermal shift assay.

The screening of the same fragment library in presence and absence of ADP resulted in identification of different fragment molecules that gave positive shifts. However, one fragmen, 5-Aminoindole, was found to be a common hit from both the screening types. The ΔTm was observed to be 1.38˚C when type-1 screening method was used. In presence of ADP in the type-2 screening method, the ΔTm was slightly increased to 2.5˚C.

2.5.11.2 Inhibition of Trehalose-6-Phosphate Synthase Activity of *M. thermoresistibile* OtsA

To test the *in vitro* effects of the hits identified on the activity of the protein, an inhibition screen was performed. Dose response curves were produced for the top hits from both the screening types, and half maximal inhibitory concentrations (IC_{50}) were determined **(Table 2.6)**. Representative dose-response curves are shown in **Figure 2.31**.

Fragment Number	IC50 Value (µM)
F2 (Type-2)	500
F5 (Type-2)	1000
F4 (Type-2)	650
F1 (Type-2)	800
F7 (Type-1)	624
F8 (Type-1)	870
F9 (Type-1)	650
F1 (Type-1)	720
F4 (Type-1)	530

Table 2.6 Summary of IC_{50} values of top hits against *M. thermoresistibile* OtsA.

Fragment hits tested from both type-1 and type-2 screening by thermal shift assay.

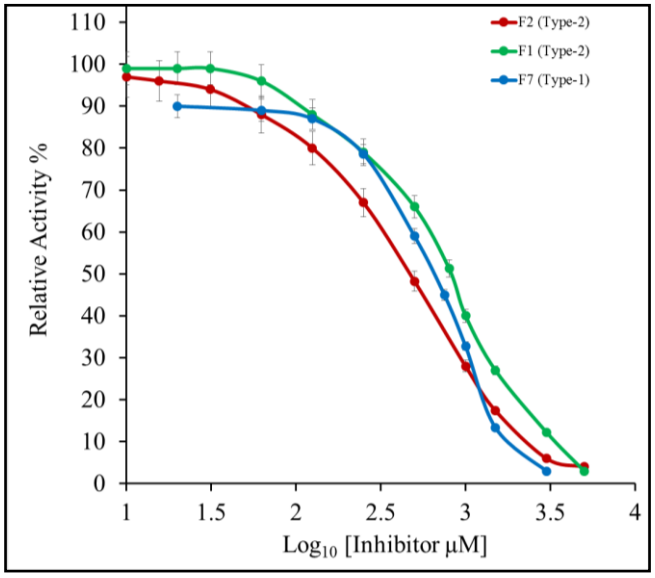


FIGURE 2.31 Inhibition of trehalose-6-phosphate synthase activity of *M. thermoresistibile* OtsA by fragments.

Representative examples of dose-response curves of F7 from type-1 screen, and F1 and F2 from type-2 screen of thermal shift assay.

2.5.11.3 Fragment-Bound Structures

M. thermoresistibile OtsA crystals were soaked or the protein was co-crystallised along with the fragments, in order to obtain fragment-complexed structures of the enzyme. However, no attemps were successful.

2.6 Discussion

Trehalose plays a pivotal role in mycobacteria and is a constituent of numerous structural components of the mycobacterial cell wall. The reactions involved in synthesis of trehalose-6-phosphate and trehalose are essential in the physiology of the bacteria. In this chapter, I describe the successful structural and functional characterisation of the mycobacterial OtsA enzyme. The analysis and alignments reported in this chapter show that OtsA from *M. tuberculosis* has 80% sequence identity with its orthologue from *M. thermoresistibile*, with full conservation of the catalytic site, which makes it an ideal surrogate model for structural and functional investigation. The recombinant enzyme from *M. thermoresistibile* was expressed reasonably in the soluble fraction, however, a mild detergent treatment was required to extract the protein from pellet in order to improve the yield. There was no apparent precipitation and instability of the protein, and the CD spectra demonstrated that protein was folded and thermally stable. The protein was functionally active and maintained its structural integrity.

The structures of *M. thermoresistibile* OtsA exhibit the typical GT-B fold of glycosyltransferases with two core Rossmann-fold domains. The overall structural fold is surprisingly identical to OtsA enzymes from *E. coli* and *B. xenovorans*, despite their low sequence identity, and also to VldE and ValL, the pseudo-glycosyltransferases involved in validamycin-A synthesis (Cavalier *et al*., 2012; Gibson *et al*., 2002a; Zheng *et al*., 2012). A significant difference lies in the catalytically important loop between Arg18 and Gly39 of *M. thermoresistibile* OtsA. It has been reported that the catalytic site loop of *E. coli* OtsA does not contain any insertion and represent the shortest loop in GT-20 family of glycosyltransferases, as compared to other members, wherein there are insertions of 8-9 residues (Gibson *et al*., 2004). Mycobacterial OtsA also has an insertion of 8-residues in this loop. However, the electron density was not visible for this insertion in any of the structures described in this work, which indicates that it is highly flexible.

There is a significant difference between *M. thermoresistibile* OtsA structures in ‘open’ conformation (observed for apo, donor-substrate bound and products-bound structures) and the ADP-G6P complex structure in ‘closed’ conformation. In *M. thermoresistibile* OtsA:ADP:G6P ternary structure, the acceptor binding site undergoes a structural and conformational change upon ligand-binding. There is a cross talk due to interactions of

ligands with the residues of the adjacent domain, which mediates movement of α1 helix and closes the catalytic site. Thereby, the enzyme adopts a 'closed' conformation that is important for its catalytic activity. Also, the catalytically important loop between Arg18 and Gly39, which is adjacent to α1 helix, forms a more organised β-hairpin motif in this ternary complex. It is possible that this loop folds differently in the apo-form, and the binding of ligands in the catalytic site drives it to fold in an ordered fashion. The substrate-gated conformational and structural changes has also been observed for other GT-B glycosyltransferases, such as in UDP:*E. coli* OtsA, Trehalose:ValL and UDP:T4 phage β-glucosyltransferase complexes (Gibson *et al*., 2004; Moréra *et al*., 2001; Zheng *et al*., 2012).

OtsA enzyme from different organisms can utilise multiple nucleotide donors, with different proclivity (Gibson *et al*., 2002b; Killick, 1979; Pan *et al*., 2002; Silva *et al*., 2005). To exemplify, *E. coli* OtsA prefers UDP-glucose (Gibson *et al*., 2002b) whereas the *M. tb* OtsA has high specificity for ADP-glucose (Pan *et al*., 2002). Utilisation of UDP-glucose by *M. smegmatis* OtsA is greatly stimulated by heparin (Lapp *et al*., 1971). In this study, I demonstrated that *M. thermoresistibile* OtsA, like the enzymes from *M. tuberculosis* and *M. smegmatis*, is capable of using different glucose-donors, but has a strong preference for ADP-glucose. This is a conclusive evidence that mycobacterial OtsA catalyses the production of trehalose-6-phosphate from any nucleotide diphosphoglucose derivatives. The donor-substrate bound structures of *M. thermoresistibile* OtsA provide a platform to understand the possible structural mechanistics that underlay substrate preferences. The preference for ADP-glucose by *M. thermoresistibile* OtsA can be attributed to its nucleotide-binding pocket, which is enlarged and deepened, as compared to *E. coli* OtsA, due to substitution of residues in the hydrophobic core of the catalytic site (particularly, Ile295 and Phe339 in *E. coli* for Leu319 and Val363 in *M. thermoresistibile* OtsA). This allows the adenine moiety to fit perfectly, with its primary amine occupying the deep, buried spot in the catalytic pocket of *M. thermoresistibile* OtsA. Although Gibson and colleagues proposed that it is possible to model GDP is place of UDP in *E. coli* OtsA structure (Gibson *et al*., 2002a), the structure of *M. thermoresistibile* OtsA bound with GDP-glucose shows that it is not feasible for the guanine moiety to occupy the deep buried position, like the adenine moiety, due to steric clashes. This was positively confirmed by the biochemical data, which demonstrated GDP-glucose was the second preferred substrate after ADP-glucose, and the thermodynamics of binding determined by ITC. The latter showed that in the tested conditions, heat of binding was observed only for ADP-glucose and not for any other NDP-glucose donor (GDP- or

UDP-glucose). Furthermore, conservation analysis also revealed that residues of the binding site interacting primarily with the nitrogenous base of the NDP-glucose donor are not conserved. This further support the possibility of diverse substrate preferences contributed due to amino-acid substitutions.

The enzyme activity was independent of the divalent ions and the pH profile showed that the maximal activity was between the pH 7.0-8.5. The enzyme also exhibited a feedback mechanism, in which the activity of the recombinant TPS was reduced not only by trehalose-6-phosphate, but also by trehalose to a relatively greater extent.

The structural-functional analyses in this chapter demonstrate the suitability of *M. thermoresistibile* as a representative system. *E. coli* OtsA does not exhibit the properties necessary for a useful surrogate for TB drug discovery because it shares low sequence identity with *M. tuberculosis* OtsA (34%) and there are differences in regions important for the enzyme mechanism (Gibson *et al*., 2002b, 2004). Thus, the definition of a structure that is more closely similar to the *M. tb* enzyme was an essential pre-requisite for structure-guided drug discovery.

This chapter also reports on the preliminary data to probe the druggability of *M. thermoresistibile* OtsA, for identification of hit fragments that could then be developed into specific inhibitors. The initial screening of a fragment library was done using a high-throughput thermal shift assay. This resulted in identification of 15 fragments that bind in the absence of ADP and 7 fragments that bind in the presence of ADP, and have a stabilisation effect, as indicated by the increase in the melting point of the enzyme upon the binding of the fragments. Attempts were made to obtain structures of the enzyme bound with hits, but these were not fruitful. The hits were validated by an *in vitro* inhibition assay. Top fragments hits were shown to have an inhibitory effect the trehalose-6-phosphate synthase activity of the enzyme, with an IC_{50} value in the range of 0.5-1 mM. This serves as the start-point for a potential drug-discovery campaign against mycobacterial OtsA.

CHAPTER-3

Bioinformatic Analysis, Expression, Purification and Crystallisation of Rv3030

3.1 Summary

Another target that was investigated in this study was Rv3030, a SAM-dependent methyltransferase in MGLP biosynthesis. The primary aim for this target was to define the structure of this enzyme, as there are no available structures of Rv3030 or its closely related orthologues in the database. This chapter describes the initial analyses of the amino-acid sequences of the protein and its orthologues in order to understand the residue conservation and the domain architecture, and a homology model of the protein structure. The design of constructs of full-length and truncated proteins and purification of proteins is also described, followed by crystallisation trials. Notwithstanding, crystallisation proved challenging. The success rate of obtaining well diffracting crystals suitable for structure determination studies was very poor, even though many trials were carried out under different conditions, with full-length and truncated proteins. Only one condition yielded crystals of the full-length orthologous protein from *M. smegmatis*, which took a very long time to grow, diffracted only to very low resolution and had low reproducibility. The understanding of structural properties of this 6-O-methyltansferase was complemented by circular dichroism, and analytical ultracentrifugation, which indicated that protein was folded, monomeric and globular in nature.

3.2 Introduction

S-adenosyl-L-methionine (also known as SAM or AdoMet) is a ubiquitous nucleoside involved in many biochemical processes such as transmethylation, transsulphuration, and aminopropylation (Parveen and Cornell, 2011). SAM serves as an important co-factor for methylation of biologically active molecules, such as proteins, lipids, polysaccharides, DNA and RNA that is catalysed by SAM-dependent methyltransferases (SAM-MTs), and S-adenosyl-L-homocysteine (SAH) is yielded as the byproduct (Kozbial and Mushegian, 2005). Methylations at oxygen (O-methylation), carbon (C-methylation) and nitrogen (N-methylation) are vital in all organisms. SAM-MTs play a pivitol role in many cellular processes, such as biosynthesis, protein repair, detoxification, signal transduction, metabolism, chromatin regulation, and nucleic acid processing (Martin and McMillan, 2002; Schubert *et al*., 2003). Five different structural classes that have been identified for SAM-

MTs based on the structure of their catalytic domains are the Rossmann-fold (Class-I), TIM barrel (Class-II), tetrapyrrole methylase (Class-III), SPOUT (Class-IV) and SET domain (Class-V) (Schubert *et al*., 2003).

The majority of SAM-MTs belong to Class-I, the typical α/β Rossmann fold. In this arrangement, there are seven β-strands arranged in the order 3214576, forming a planar, central β-sheet **(Figure 3.1)**. The β-strands alternate with six α-helices in the polypeptide, three arranged on each side of the β-sheet (Martin and McMillan, 2002). The N-terminal β-strand is in the middle of the sheet and the seventh strand is antiparallel to all other strands (Kozbial and Mushegian, 2005). Most of SAM-MTs comply with this design, although occasionally the addition of an extra helix or β-hairpin, or deletion of one or both of β-strands 6 and 7 are observed (Bujnicki, 2000; Martin and McMillan, 2002).

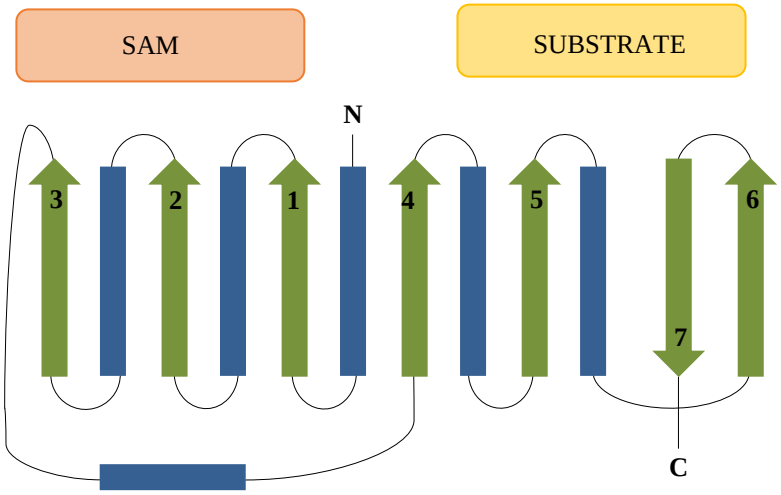


FIGURE 3.1 Schematic representation of the Rossmann fold topology of SAM-MTs.

Helices are shown as blue rectangles, strands as green arrows. SAM and substrate binding regions are shown. (Figure adapted from Martin and McMillan, 2002).

The residues of the SAM-binding site are not conserved, but the co-factor occupies the same position of the structural fold in the enzyme and adopts same conformation (Martin and McMillan, 2002; Schubert *et al*., 2003). The SAM-binding region is formed by strands 1, 2 and 3, the N-terminal part of the β-sheet, and there are only two conserved motifs in the

SAM-binding site of Class-I SAM-MTs (Martin and McMillan, 2002). Motif-I is the consensus sequence 'GxGxG' at the end of β1 strand that is considered to be characteristic of the SAM-binding site and is used to bind the adenosyl moiety of SAM (Schubert *et al*., 2003). Motif II is a strongly conserved acidic residue present at the C-terminus of β2 strand that forms hydrogen bonds with both the ribose hydroxyls of SAM (Schubert *et al*., 2003). The substrate-binding region is formed by the C-terminal part of the β-sheet (Martin and McMillan, 2002).

Methyltransferases of a particular structural class share high similarity of the structural fold, but it does not indicate conservation of the amino-acid sequence, and the sequence identity can be as low as 10% (Cheng and Roberts, 2001). Nevertheless, the mode of catalysis is conserved and methyltransferases catalyse direct transfer of methyl group to the substrate via an S_N2-substitution mechanism (Horowitz *et al*., 2013).

M. tb proteome has 121 methyltranferases that are involved in an array of cellular process (Grover *et al*., 2016). Of these, Rv3030 is an essential SAM-dependent methyltransferase involved in MGLP biosynthesis in *M. tb* that carries out methylation of glucoses of the MGLP chain at 6-O position. 6-O-methyl-D-glucose was first identified in *Mycobacterium phlei* (Lee and Ballou, 1964). The methyltransferase activity was described in *M. phlei* by Ferguson and Ballou, which was shown to transfer methyl groups from SAM to 6-O position of D-glucose in the oligosaccharides (Ferguson and Ballou, 1970). Later, a study by Grellert and Ballou demonstrated that for the methyltransferase activity to occur, the substrate needs to be partially acetylated, and provided evidence that the degree of acylation controls the positioning of O-methylation (Grellert and Ballou, 1972).

3.3 Basic Principles of Techniques Deployed

3.3.1 Sedimentation-Velocity Analytical Ultra-Centrifugation (SV-AUC)

This technique is used for quantitatively characterising the molecules in a solution. It provides useful information about sample homogeneity, size, hydrodynamic shape of molecules and association state, using first principles (Schuck, 2016).

3.3.2 Ligand-Based Nuclear Magnetic Resonance

Ligand based NMR techniques, saturation-transfer difference (STD) and Carr-Purcell-Meiboom-Gill (CPMG), are important for screening ligands or small-chemical moieties that bind to protein. Signal intensities are observed for free and ligand-bound states. STD-NMR is based upon nuclear Overhauser effect (NOE), in which the saturation is transferred from protein to the bound ligand through spin diffusion. Two spectra are recorded, an 'on-resonance' spectrum in which there is selective protein saturation and the magnetisation transfer to the bound ligand is measured, and an 'off-resonance' spectrum in which there is no protein saturation and thus, no magnetisation is transferred to the ligand. A difference spectrum is produced and intensities are observed only for the ligands that receive magnetisation transfer from the protein (Meyer *et al*., 2004; Viegas *et al*., 2011). A CPMG experiment is based on the relaxation properties of the molecules. Protein molecules have slow relaxation rate whereas ligands have faster relaxation times. When a ligand binds to the protein, there is a decrease in intensity of the measured signal because of the decrease of free molecules in solution that contribute to signal intensity.

3.4 Methodology

3.4.1 Bioinformatic Analysis of the Primary Sequence, Sequence Alignment and Prediction of Secondary Structures

The full-length protein sequence of Rv3030 from *M. tuberculosis* was obtained from the TubercuList database (http://tuberculist.epfl.ch/) (Lew *et al*., 2011), and the sequence of its orthologue in *M. smegmatis* (*MSMEG_2350*) from the SmegmaList database (http:// mycobrowser.epfl.ch/smegmalist.html) (Kapopoulou *et al.*, 2011) by searching with the gene name. The orthologous proteins in *M. thermoresistibile, M. leprae, M. marinum, M. abscessus, M. ulcerans and N. farcinica* were identified by performing a protein-BLAST (blastp) search (http://blast.ncbi.nlm.nih.gov/Blast.cgi) (Altschul *et al*., 1990) with the *M. tuberculosis* protein sequence. The amino-acid sequences were obtained from National Centre for Biotechnology Information (NCBI) database (http://www.ncbi.nlm.nih.gov), and sequences were aligned using Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) (Goujon *et al*., 2010; Sievers *et al*., 2014). The alignment was analysed and edited using Jalview (Waterhouse *et al*., 2009). The methyltransferase family and domain alignment coordinates were identified with Pfam (http://pfam.sanger.ac.uk) (Punta *et al*., 2012). The secondary structures of the protein were predicted using PSIpred (http://bioinf.cs.ucl.ac.uk/psipred/) (Jones, 1999).

3.4.2 Protein Homology Modelling of Full-Length Orthologue of Rv3030 from *M. smegmatis*

A homology model of full-length protein of orthologue of Rv3030 in *M. smegmatis* was built using ModSuite (http://mordred.bioc.cam.ac.uk/modsuite), which encompasses BATON (based on Comparer), FUGUEALI and MODELLER (developed by Dr. Bernardo Ochoa, currently for the internal use of Blundell group members only). The templates for model building were identified using FUGUE (http://tardis.nibio.go.jp/fugue/prfsearch.html) (Shi *et al*., 2001) and the NCBI protein-BLAST. The protein sequences and PDB files of these templates were collected from the RCSB-Protein Data Bank (PDB) (http://www.rcsb.org) (Rose *et al*., 2013) by searching with their respective PDB IDs. The templates were first

structurally aligned to each other with help of BATON, and then a sequence-structure alignment of the query protein sequence with previously aligned templates was done with FUGUEALI. The final model of the protein was generated using MODELLER (Sali and Blundell, 1993). Dr. David Ascher and Dr. Bernardo Ochoa helped in *in silico* modelling of the protein structure.

3.4.3 DNA, Bacterial Strains, Cloning and Expression Trials of Rv3030 and its Orthologues

Genomic DNA from *M. tuberculosis* H37Rv was obtained from ATCC (ATCC25618D-2), *M. thermoresistibile* (DSMZ 44167) DNA was obtained from the Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ, Braunschweig, Germany), and the genomic DNA from *M. smegmatis* strain mc^2 155 was obtained from Dr. Nuno Empadinhas (University of Coimbra, Portugal). *E. coli* DH5α (Invitrogen) and BL21 (DE3) (New England Biolabs) strains were used for cloning and protein expression, respectively.

Dr. Vitor Mendes did the cloning and expression trials of full-length Rv3030 and its orthologues from *M. smegmatis* and *M. thermoresistibile*, and provided the BL21 glycerol cell stock of the expression clone. Rv3030 and its orthologue in *M. smegmatis* were cloned in BamHI and HindIII restriction sites in pET-28a SUMO vector. Rv3030 orthologue in *M. thermoresistibile* was cloned in NcoI and EcoRI restriction sites in pHAT-5 vector, containing a C-terminal, non-cleavable His_6 tag and ampicillin resistance marker (obtained from Dr. Hyvönen's group, Department of Biochemistry, University of Cambridge) (Peränen *et al*., 1996).

Genes with truncations of varying lengths at N-terminus of full-length orthologue of Rv3030 from *M. smegmatis*, were cloned into BamHI and HindIII restriction sites in pET-28a SUMO vector. Details of constructs are given in **Table 3.1**.

CONSTRUCT	DESCRIPTION	PRIMER SEQUENCE (5'-3')
-4N 3030Msmg pET-28a SUMO	4 residues truncated from the N-terminus	Forward- ATAGGATCCGATAACGCCCTTCCTTCGGC Reverse- ATTAAGCTTTCATGGCCGCACTGCTATCG
-8N 3030Msmg pET-28a SUMO	8 residues truncated from the N-terminus	Forward- ATAGGATCCCCTTCGGCTTTACCCCTCAC Reverse- ATTAAGCTTTCATGGCCGCACTGCTATCG
-14N 3030Msmg pET-28a SUMO	14 residues truncated from the N-terminus	Forward- ATAGGATCCACCGGCGAGCGGACCATCCC Reverse- ATTAAGCTTTCATGGCCGCACTGCTATCGCC
-15N 3030Msmg pET-28a SUMO	15 residues truncated from the N-terminus	Forward- ATAGGATCCGGCGAGCGGACCATCCCGG Reverse- ATTAAGCTTTCATGGCCGCACTGCTATCGCC
-17N 3030Msmg pET-28a SUMO	17 residues truncated from the N-terminus	Forward- ATAGGATCCCGGACCATCCCGGGCCTGG Reverse- ATTAAGCTTTCATGGCCGCACTGCTATCGCC
-24N 3030Msmg pET-28a SUMO	24 residues truncated from the N-terminus	Forward- ATAGGATCCGAGGAGAACTACTGGTTCCG Reverse- ATTAAGCTTTCATGGCCGCACTGCTATCG

TABLE 3.1 Details of constructs and primers used for cloning and expression of truncated orthologues of Rv3030 from *M. smegmatis.*

3030Msmg refers to the orthologue of Rv3030 from *M. smegmatis.* Restriction enzymes (BamHI in forward primer and HindIII in reverse primer) are underlined.

PCR was done as described in the section 2.4.3, but with 15 sec annealing step at 58°C, and 20 sec extension step at 70°C. The amplicons generated and the vector were digested with the BamHI and HindIII restriction enzymes (ThermoScientific), and ligated using T4 DNA ligase (New England Biolabs) at room temperature for 10 min. The ligation products were transformed in *E. coli* DH5α competent cells, plated on LB agar-kanamycin plates, and incubated overnight at 37°C. Single colonies were randomly picked and inoculated in LB media (Melford) with 30 μg/ml kanamycin (Sigma), and grown overnight at 37°C, with constant shaking at 220 rpm. Plasmids were isolated and purified according to the manufacturer's protocol (ThermoScientific GeneJet Plasmid Miniprep Kit). The integrity of clones was confirmed by sequencing (DNA Sequencing Facility, Department of Biochemistry, University of Cambridge, UK), and plasmids were transformed into *E. coli* BL21 (DE3) strain.

Expression trials of truncated constructs were done as described in the section 2.4.4.1, but kanamycin was used as the selection marker at final concentration of 30 μg/ml. The proteins' identity was confirmed by MALDI fingerprinting (PNAC facility, Department of Biochemistry, University of Cambridge, UK).

3.4.4 Expression and Purification of Recombinant Orthologues of Rv3030

3.4.4.1 Expression and Purification of Recombinant Full-Length Orthologue of Rv3030 from *M. smegmatis*

The recombinant orthologous protein of Rv3030 from *M. smegmatis*, with a cleavable, N-terminal His_6-SUMO tag was purified from the soluble fraction.

Culture growth and expression: A loop of frozen BL21 cells of the expression clone was plated on LB-agar-kanamycin plate and incubated overnight at 37°C. All the cells were collected from the plate and inoculated in 100 ml LB media with 30 μg/ml kanamycin. The culture was grown overnight at 37°C, with constant shaking at 220 rpm. The culture was scaled up to 2 litres by adding 30 ml of the overnight culture in each 1-litre fresh LB media with kanamycin. Cells were grown at 37°C till OD_{600} reached 0.8, cooled down to 20°C, and induced with 0.5 mM IPTG for protein expression, and grown overnight (14-16 hours) at 20°C with constant shaking.

Cell harvesting and lysis: Cells were harvested by centrifugation (4200 x g, 30 min, 4°C) and the pellet was suspended in buffer (50 mM Tris-HCl pH 7.5, 20 mM Imidazole, 500 mM NaCl, cocktail of protease inhibitor (Roche)) and, 10 μg/mL DNase-I (Sigma) and 5 mM $MgCl_2$ were added to the cell suspension. The cells were lysed by passing them through Emulsiflex (GlenCreston) four times, followed by centrifugation (23,500 x g, 30 min, 4°C) to remove cell debris. The supernatant was collected and filtered using 0.2 μm syringe filter (Sartorius Stedim).

Affinity purification: A pre-packed sepharose column (HiTrap IMAC FF, 5 ml, GE Healthcare) was charged with 0.1 M $NiSO_4$, according to the manufacturer's protocol, and equilibrated with purification buffers. Supernatant was bound to the column with the help of

a peristaltic pump at room temperature. The column was washed for 30 column volume (CV) with wash buffer (50 mM Tris-HCl pH 7.5, 20 mM Imidazole, 500 mM NaCl), and 20 ml protein was eluted with elution buffer (50 mM Tris-HCl pH 7.5, 500 mM Imidazole, 500 mM NaCl). The purity of the protein was determined by SDS-PAGE.

Cleavage of tag, dialysis and reverse binding: The SUMO tag was subsequently cleaved with Ulp1 protease, after affinity purification. Ulp1 protease was added to the eluted protein at a ratio of 1 mg Ulp1 protease to 200 mg protein, to cleave the His_6-SUMO tag. The protein-protease mixture was then dialysed in the dialysis buffer (20 mM Tris-HCl pH 7.5, 20 mM Imidazole and 500 mM NaCl), using 10 kDa MWCO dialysis membrane (Spectrum Labs), and incubated overnight at 4°C. The digested protein was bound to the Ni-sepharose column to separate the tag-less protein from the tag. This is known as reverse binding, where the His_6-SUMO tag and Ulp1 protease bound to Ni-sepharose column while the tag-less target protein flowed through. Flow through containing target protein was collected and concentrated by centrifuging (4500 x g, 30 min, 4°C) in 10 kDa MWCO concentrators.

Gel filtration: The protein was loaded onto a Superdex-200 column (HiLoad 16/600 Superdex 200 pg, GE Healthcare), pre-equilibrated with 100 mM NaCl and 20 mM Tris-HCl pH 7.5. Purification was done at room temperature. Protein was eluted in 1 ml fractions and the purity of fractions was determined by SDS-PAGE. The purest fractions were pooled and concentrated by centrifuging (4500 x g, 30 min, 4°C) in 10 kDa MWCO concentrators. The purity of the protein was assessed with SDS-PAGE and the final protein concentration was measured at 280 nm with NanoDrop spectrophotometer (Labtech, UK).

N.B. Another batch of full-length recombinant orthologous protein of Rv3030 from *M. smegmatis* was purified but the fusion tag was **not** cleaved.

3.4.4.2 Expression and Purification of Recombinant Full-length Orthologue of Rv3030 from *M. thermoresistibile*

The full-length recombinant orthologue of Rv3030 from *M. thermoresistibile*, with C-terminus, non-cleavable His_6 tag was expressed and purified from the soluble fraction. The protocol for expression and purification described in the section 3.4.4.1 was used with some

modifications. 100 µg/ml ampicillin was used as the selection antibiotic. There was no tag cleavage and the protein was dialysed overnight in 100 mM NaCl and 20 mM Tris-HCl pH 7.5 buffer. Gel filtration was done as described but at 4°C.

3.4.4.3 Expression and Purification of Recombinant Truncated Orthologues of Rv3030 from *M. smegmatis*

The truncated recombinant orthologues of Rv3030 from *M. smegmatis*, with N-terminus, cleavable His_6-SUMO tag were expressed and purified from the soluble fraction. The protocol for expression and purification was as described in the section 3.4.4.1.

3.4.5 Crystallisation of Orthologues of Rv3030

3.4.5.1 Crystallisation Screening of Full-Length Orthologue of Rv3030 from *M. thermoresistibile*

Initial crystallisation screening was performed for the full-length orthologue of Rv3030 from *M. thermoresistibile* with C-terminal His_6 tag, using a sitting-drop vapour-diffusion method at 18˚C. The screens used were: classics, pH clear-I (Qiagen), PEGs-I (Qiagen) and wizard I&II. 0.3 µl protein at concentrations of 10 mg/ml and 20 mg/ml, was respectively mixed with the well solution in two different protein: precipitant ratios of 1:1 and 1:2 using the Phoenix robot.

3.4.5.2 Crystallisation Screening of Full-Length Untagged Orthologue of Rv3030 from *M. smegmatis*

A. Initial crystallisation screening: This was done as described in section 3.4.5.1, but the protein only at concentration of 20.5 mg/ml was used.

B. Co-crystallisation with SAM and SAH: Protein (at 20 mg/ml) was incubated with 2 mM SAM and 2 mM SAH, in separate reactions, on ice for 30 minutes. Initial crystallisation screening was done as described in section 3.4.5.1.

C. Co-crystallisation with $MgCl_2$: Protein (at 10 mg/ml) was incubated with 10 mM $MgCl_2$ at room temperature for 30 minutes. The screens used were: classics, pH clear-I, pH clear-II (Qiagen), PEGs-I, PEGs-II (Qiagen), cryo (Qiagen), MPD (Qiagen) and wizard I&II. 0.2 µl protein-$MgCl_2$ mixture was mixed with the well solution in two different protein: precipitant ratios of 1:1 and 1:2 using the Phoenix robot.

D. Limited proteolysis: Limited proteolysis of the full-length protein was performed using JBS Floppy Choppy kit (Jena Biosciences). Pre-screening was done to identify the suitable protease, according to manufacture's protocol. An additional, time-dependent activity of protease was done with same protocol, except that the incubation times were 30 minutes, 1 hour and 2 hours. After determining the suitable protease, the concentration required and the time for digestion (in this case: chymotrypsin, 1 µg of protease for 100 µg protein, digestion for 2 hours); protease and protein were incubated at room temperature for 2 hours. The reaction was stopped by adding 1 mM PMSF. Initial crystallisation screening with the digested protein was done. The screens used were: classics, pH clear-I, pH clear-II, PEGs-I, PEGs-II, cryo, MPD and wizard I & II. 0.2 µl digested protein was mixed with the well solution in a ratio of 1:1 using the Phoenix robot.

E. Optimisation of phase separation condition: A phase separation condition (4M sodium formate) identified from the initial screening of full-length protein was optimised, using the hanging-drop vapour-diffusion method, as follows:

1. A grid was prepared for different concentrations of protein (12, 25, 30 and 50 mg/ml) and sodium formate (0.5, 1, 2, 3, 4 and 5 M) using a 24-well Linbro plate (Molecular Dimensions).
2. A grid was set with different concentrations of sodium formate (0.5, 1.0, 2.0, 3.0, 4.0, 5.0 M), 0.1 M Tris-HCl buffer at varying pH values (7.0, 7.5, 8.0, 8.5), but with constant NaCl concentration (0.2 M). Protein at 10 mg/ml concentration was used for crystallisation.
3. A grid was set with the same conditions in (2) except that NaCl was not added.

F. Optimization of crystallisation condition: The condition: 4 M sodium formate, 0.1 M Tris-HCl pH 8.5, 0.2 M NaCl, where crystals were formed, was further optimised using the hanging-drop vapour-diffusion method, as follows:

1. Reproducibility of the crystal condition: A grid was set with different concentrations of sodium formate (2.0, 2.5, 3.0, 3.5, 4.0, 4.5 M), 0.1 M Tris-HCl buffer at varying pH values (7.5, 8.0, 8.5, 9.0), but with constant NaCl concentration (0.2 M). Protein at 10 mg/ml concentration was used for crystallisation.
2. Streak seeding: The same condition where crystals were obtained was set up. Drops (of protein and precipitant at 1:1 ratio) were seeded with previously formed crystals using a cat whisker at different time points: immediately, 1 hour, 3 hours, 5 hours, 1 day, 2 days, 3 days and 1 week after setting drop.
3. Additive screening: Additives were added to the condition in which crystals were obtained. Additive screen HT from Hampton Research was used.
4. A grid was set with concentrations of sodium formate at narrow differences (3.90, 3.95, 4.00, 4.05, 4.10, 4.15 M), 0.1 M Tris-HCl buffer at varying pH (7.5, 8.0, 8.5, 9.0), but with constant NaCl concentration (0.2 M). Protein at 10 mg/ml concentration was used for crystallisation.
5. Matrix seeding (in the same condition where crystals were previously obtained): A seed stock of crystals was prepared by fishing out all crystals along with the mother liquor and pooling them together in a vial, containing a bead (Hampton Research). Crystals were crushed by vortexing in intervals for 15minutes. Drops (1 µl protein + 1 µl precipitant + 0.2 µl seeds) were set using Oryx robot (Douglas Instruments).
6. A grid was set with different protein concentrations (2, 3.5, 5, 6.5, 8, and 9.5 mg/ml) and different protein: precipitant ratios (1:1, 1:2, 1:3, and 1:4) in the same condition where low diffracting crystals were previously obtained.
7. Screening different salts: A grid was set with different concentrations of sodium formate (2.0, 2.5, 3.0, 3.5, 4.0, 4.5 M), different salts (sodium chloride, sodium citrate, sodium formate, sodium acetate, Na/K tartrate, ammonium acetate, ammonium chloride, sodium melonate) (0.2 M), but with 0.1 M Tris-HCl buffer pH 8.0. Protein at 10 mg/ml concentration was used for crystallisation.

G. Matrix seeding using Classics screen: The seed stock was diluted 5 times with the mother liquor, and was used for seeding and screening new crystal conditions, using the sitting-drop vapour-diffusion method.

H. Co-crystallisation with SAM and maltohexaose in crystal condition: Protein (at 5 mg/ml and 10 mg/ml) was incubated with 2 mM SAM and 2 mM maltohexaose, in separate reactions, on ice for 30 minutes before setting crystallisation in the condition: 4 M sodium formate, 0.1 M Tris-HCl pH 8.5 and 0.2 M NaCl. For each protein concentration and each ligand, different protein: precipitant ratios were set (1:1, 1:2, and 1:3). Crystallisation was performed using the hanging-drop vapour-diffusion method.

3.4.5.3 Crystallisation Screening of Full-Length Tagged Orthologue of Rv3030 from *M. smegmatis*

Initial screening trials were done with the tagged protein as stated in section 3.4.5.1, but the protein only at concentration of 12 mg/ml was used.

3.4.5.4 Crystallisation Screening of Truncated Orthologues of Rv3030 from *M. smegmatis*

A. Initial screening trials: These were done as described in section 3.4.5.1, and each of the truncated protein, at a concentration of 20.5 mg/ml, was screened.

B. Crystallisation of truncated proteins in the condition where full-length protein crystallised: A grid was set where conditions with concentrations of sodium formate at narrow differences (3.90, 3.95, 4.00, 4.05, 4.10, 4.15 M), 0.1M Tris-HCl buffer at varying pH (7.5, 8.0, 8.5, 9.0), but with constant NaCl concentration (0.2 M) were screened. Crystallisation was set for each of the truncated protein (at 10 mg/ml concentration).

C. Optimisation of microcrystalline condition for a truncated protein (17 residues truncated at N-terminus): The condition (1.8 M ammonium sulphate, 0.1 M MES pH 6.5, 0.01 M $CoCl_2$, 10 mg/ml protein) where microcrystals were formed was further optimised, using the hanging-drop vapour-diffusion method, as follows:

1. A grid was set with different concentrations of ammonium sulfate (0.9, 1.2, 1.5, 1.8, 2.1, 2.4 M), 0.1 M MES buffer at varying pH values (5.5, 6.0, 6.5), 0.1 M HEPES buffer at varying pH values (7.0, 7.5), 0.1 M Tris-HCl buffer at varying pH values (7.5, 8.0, 8.5), and with constant $CoCl_2$ concentration (0.01 M). Protein at 10 mg/ml concentration was used for crystallisation.
2. Additive screening: Additives were added to the condition in which crystals were obtained. Additive screen from Hampton Research was used.
3. A grid was set with different concentrations of ammonium sulfate (0.9, 1.2, 1.5, 1.8, 2.1, 2.4 M), and different buffers (0.1 M Na/K PO_4 pH 6.0, 0.1 M Bis-Tris-HCl pH 6.0, 0.1 M MES pH 6.5, 0.1 M MOPS pH 6.5, 0.1 M ADA pH 6.5, and 0.1 M Tris-HCl pH 7.0), and with constant $CoCl_2$ concentration (0.01 M). Protein at 10 mg/ml concentration was used for crystallisation.

D. Optimisation of phase separation condition for a truncated protein (17 residues truncated at N-terminus): The condition (1.8 M ammonium sulphate, 0.1 M MES pH 6.5, 0.01 M $CoCl_2$, 0.5% PVP-K15, 10 mg/ml protein) where there was a phase separation was further optimised, using the hanging-drop vapour-diffusion method. A grid was set with different concentrations of ammonium sulphate (0.9, 1.2, 1.5, 1.8, 2.1, 2.4 M), 0.1 M MES buffer at varying pH values (5.5, 6.0, 6.5), 0.1 M HEPES buffer at varying pH values (7.0, 7.5), 0.1 M Tris-HCl buffer at varying pH values (7.5, 8.0, 8.5), and with constant $CoCl_2$ concentration (0.01 M) and PVP-K15 (0.5%). Protein at 10 mg/ml concentration was used for crystallisation.

3.4.6 X-ray Data Collection and Processing

The crystals of full-length orthologue of Rv3030 from *M. smegmatis* were first checked in-house with the X-ray diffractometer (ICARUS). Crystals were cryoprotected by soaking them briefly in a solution containing the mother liquor and 15% MPD, and were flash frozen in liquid nitrogen for data collection. An X-ray diffraction dataset of the crystal was collected at I03 beamline of Diamond Light Source synchrotron (DLS, Oxford, UK). Data were processed with MOSFLM (MRC, Laboratory of Molecular Biology, Cambridge, UK) (Battye *et al*., 2011).

3.4.7 Circular Dichroism of Orthologue of Rv3030 from *M. smegmatis*

The full-length orthologue of Rv3030 from *M. smegmatis* at 0.25 mg/ml, in a buffer of 20 mM sodium phosphate pH 7.5 and 50 mM NaF, was placed in quartz cuvette with 1 mm path-length (Hellma). A wavelength scan was recorded for the sample on the Aviv model 410 spectropolarimeter from 260 nm to 190 nm, with 1 nm increments and 3 sec averaging time. A control run was performed with the buffer. For the buffer, the raw data from three measurements were averaged. For the protein sample, the raw data from three measurements were averaged, corrected for the buffer signal, and smoothed. The CD signal was reported in theta machine units (in millidegrees). It was converted into mean residual ellipticity (MRE [θ], in degree cm^2 $dmol^{-1}$ $residue^{-1}$), using the following equations:

$$MRE = (MRW \times millidegrees) / 10 \times d \times c$$

where, d- path length of the cuvette in cm
c- concentration in mg/ml
MRW- mean residual weight

$$MRW = molecular\ mass / (N - 1)$$

where, N- number of amino acids

The data was deconvoluted using CDSSTR programme on the DichroWeb (http://dichroweb.cryst.bbk.ac.uk/) (Whitmore and Wallace, 2004).

3.4.8 Sedimentation Velocity-Analytical Ultracentrifugation of Orthologue of Rv3030 from *M. smegmatis*

The full-length protein was dialysed overnight in 50 mM NaCl, 20 mM sodium phosphate pH 6.5 buffer. Sedimentation velocity-analytical ultracentrifugation (SV-AUC) was performed with an XL-1 analytical ultracentrifuge (Beckman Optima), with double-sector charcoal filled EPON cells, and AN60Ti rotor. The concentration of protein used was 1.2 mg/ml. The buffer

in which protein was dialysed was used as the reference. The samples were spun at a speed of 50,000 g at 20˚C. Independent measurements were made using both the UV absorbance and interference optical systems. A total of 300 scans were recorded at 280 nm. The data were analysed using SEDFIT software (Schuck, 2000).

3.4.9 Ligand-Based NMR

The ^{1}H-CPMG experiment spectrum was recorded at 298 K on Bruker AvanceIII AV600 spectrometer equipped with HCNF probe and actively shielded *z*-gradients. Samples (600 µl) containing 1 mM SAM and 10% D_2O, in the presence and absence of 10 µM protein, were prepared in 50 mM NaCl and 20 mM sodium phosphate pH 6.5 buffer. Samples were placed in 3 mm NMR tubes for data acquisition. The resulting spectra were analysed with Bruker TopSpin software.

3.4.10 Isothermal Titration Calorimetry

The thermodynamics of the binding between the full-length orthologue of Rv3030 from *M. smegmatis* and its co-factor, SAM, was characterised using a Microcal ITC200 titration calorimeter (Microcal) at 25 ºC. Protein at concentration of 50 µM in 20 mM Tris-HCl pH 7.5, 100 mM NaCl was used for all titrations. 0.5 mM co-factors were injected in 2 µl aliquots. A total of 18 injections with 140 sec spacing was done. The injections were further continued to achieve saturation, by re-filling the syringe with the titrant, while keeping the cell untouched. Data were concatenated and analysed by fitting a simple single-site model using Origin software (Microcal).

3.5 Results

3.5.1 Sequence Alignment and Analysis of Rv3030 and its Orthologues

MGLPs are found in mycobacteria and closely related species in Nocardia and Corynebacterium. The BLAST analysis of *M. tuberculosis* Rv3030 protein sequence shows that it has orthologues in closely related species in Mycobacteria and Nocardia, including *M. thermoresistibile* (69% identity), *M. smegmatis* (MSMEG_2350 (74% identity), *M. leprae* (77% identity), *M. marinum* (69% identity), *M abscessus* (70% identity), *M. ulcerans* (79% identity) *and N. farcinica* (64% identity). The sequences were aligned using Clustal Omega, and the predicted secondary structure elements of Rv3030 orthologue from *M. smegmatis* were mapped on the alignment **(Figure 3.2)**.

The sequence alignment of Rv3030 with its orthologues showed that protein is well conserved in the mycobacterial and the closely related Nocardia species. The residues in the methyltransferase domain are highly conserved and span about one-third of the protein. The alignment coordinates of this domain in Rv3030 and its orthologues from *M. smegmatis* and *M. thermoresistibile,* as obtained from the Pfam database, are given in **Table 3.2**. The well-conserved acidic residue, glutamic acid, and the consensus sequence GxGxG (x is any residue) of motif-I, and the conserved acidic residue of motif-II are pesent.

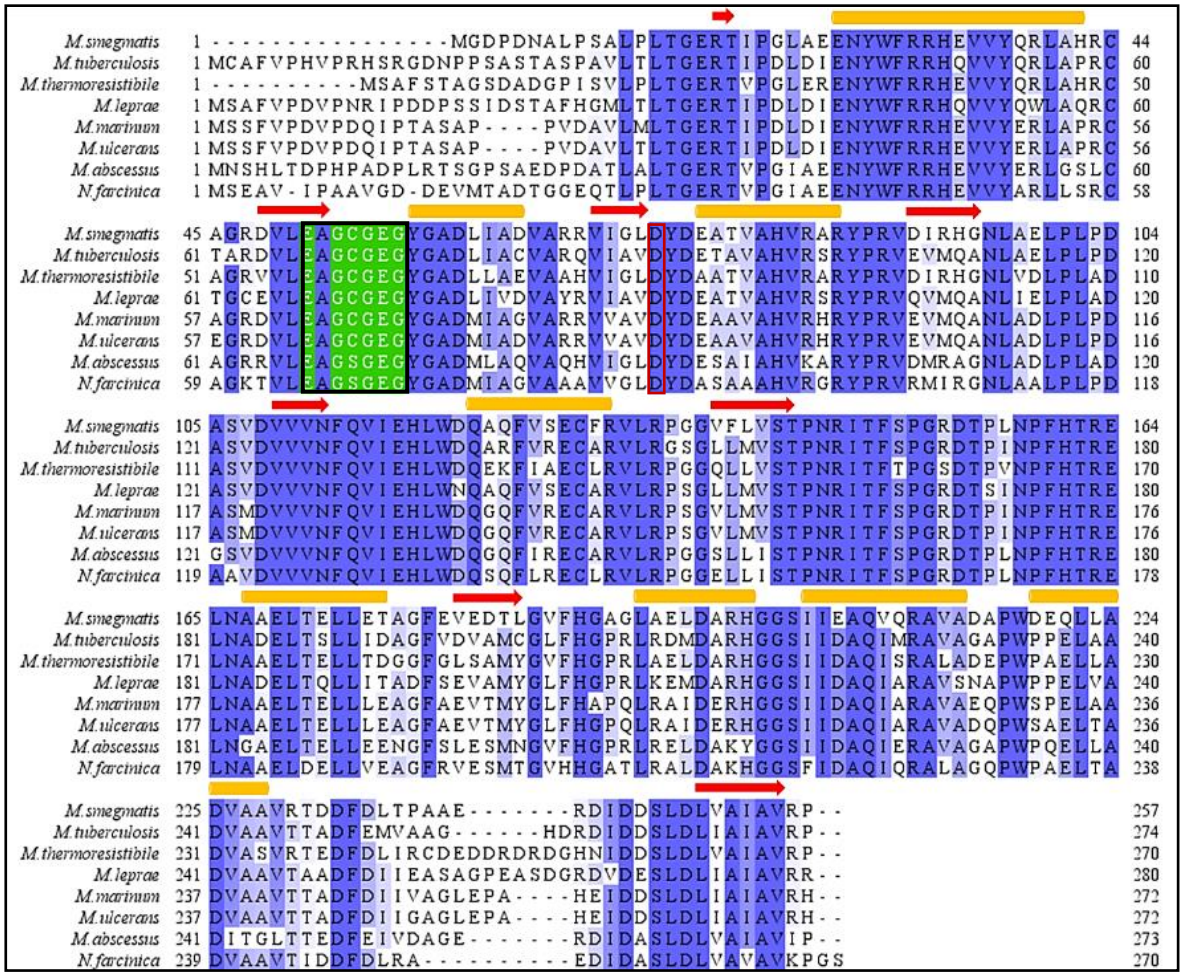


FIGURE 3.2 Sequence alignment of Rv3030 and its orthologues.

The sequence alignment, produced by Clustal Omega, shows that the protein is highly conserved between all the species. Predicted secondary structural features are illustrated as- α- helix: yellow tube; β-strand: red arrow. Motif-I of SAM-MTs, containing the well-conserved acidic residue (glutamic acid) and the consensus sequence "GxGxG", is marked with a black box. Motif-II, the conserved acidic residue Asp (D), is marked with a maroon box. Alignment coloured according to conservation.

PROTEIN	METHYLTRANSFERASE DOMAIN ALIGNMENT COORDINATES (RESIDUE RANGE)
Rv3030	66-157
Rv3030 orthologue from *M. smegmatis*	50-141
Rv3030 orthologue from *M. thermoresistibile*	56-147

TABLE 3.2 Alignment coordinates of the methyltransferase domain in Rv3030 and its orthologues.

As given in the Pfam database.

3.5.2 Homology Model of Full-Length Orthologue of Rv3030 from *M. smegmatis*

A homology model was generated for the full-length, untagged orthologue of Rv3030 from *M. smegmatis*. This helped in understanding the overall structural properties of the protein, and formed the basis for preparation of different constructs for expression, purification and structure determination. The templates used for building model were methyltransferase proteins with structures available in PDB (PDB ID of templates-2GS9, 3EGE, 1VL5, 3CCF, 3CC8) **(Table 3.3)**.

PDB ID	PROTEIN	ORGANISM
2GS9_A	Hypothetical protein TT1324	*Thermus thermophilus HB8*
3EGE_A	Putative methyltransferase from antibiotic biosynthesis pathway (YP_324569.1)	*Anabaena variabilis ATCC 29413*
1VL5_B	Putative methyltransferase (BH2331)	*Bacillus halodurans C-125*
3CCF_B	Putative methyltransferase (YP_321342.1)	*Anabaena variabilis ATCC 29413*
3CC8_A	Putative methyltransferase (bce_1332)	*Bacillus cereus ATCC 10987*

TABLE 3.3 Structural templates used for homology modelling of orthologue of Rv3030 from *M. smegmatis.*

The model showed that the protein had one domain, with centre of it spanned by a β-sheet, typical of the Rossmann fold topology of SAM-MTs. An important feature observed was that a few residues at the N-terminus of the protein were unstructured **(Figure 3.3)**.

Solvent inaccessible	UPPER CASE	X	hydrogen bond to main chain amide	**bold**	**x**
Solvent accessible	lowercase	x	hydrogen bond to main chain carbonyl	underline	x
α-helix	red	x	3_{10} helix	maroon	x
β-strand	blue	x	positive phi	*italic*	*x*

A.

B.

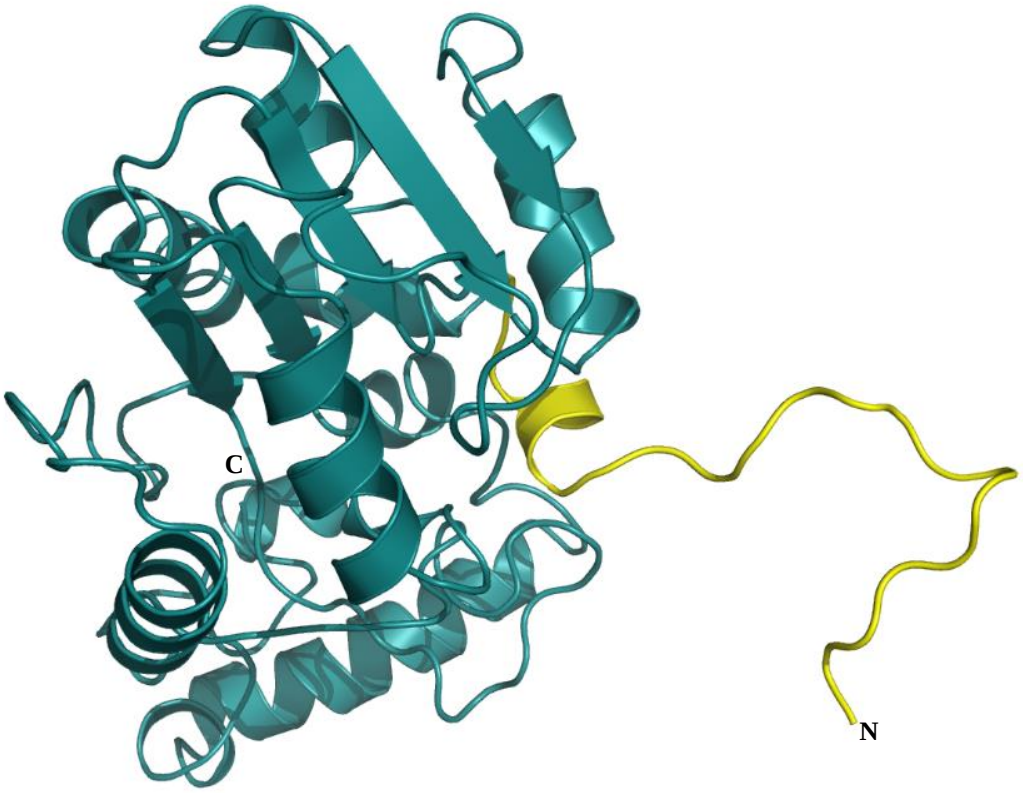


FIGURE 3.3 Homology model of three-dimensional structure of full-length orthologue of Rv3030 from *M. smegmatis.*

A. Structural alignment of sequences of target protein and templates (produced by FUGUEALI). A part of N-terminus of target protein (highlighted in yellow) does not align to any secondary structure element. The methyltransferase domain of all the proteins aligns fairly well. (3030M.smg refers to Rv3030 orthologue from *M. smegmatis*).

B. Model of the target protein. Unstructured strand at the N-terminus of protein (corresponding to the yellow highlighted part in the structural alignment above) is coloured yellow.

3.5.3 Full-length Orthologue of Rv3030 from *M. thermoresistibile*

3.5.3.1 Expression and Purification

The expression trials indicated that SUMO-tagged *M. tuberculosis* protein was insoluble and found in inclusion bodies (Dr. Vitor Mendes, personal communication). Thus, orthologues from *M. thermoresistibile* and *M. smegmatis* were expressed and purified.

The expression of the orthologous *M. thermoresistibile* gene in *E. coli* resulted in high-level production of recombinant protein fused with non-cleavable His_6 tag at its C-terminus. A strong band was seen at 33 kDa. The gel filtration profile showed that the recombinant His_6-tagged protein behaved as a monomer in solution. The protein obtained was pure enough, as seen on the gel, for crystallisation experiments **(Figure 3.4)**.

3.5.3.2 Crystallisation Trials

The tagged *M. thermoresistibile* orthologous protein was screened against approximately 1600 crystallisation conditions at two different concentrations of protein, 10 mg/ml and 20 mg/ml. Heavy precipitation was observed in the majority of conditions, with clear drops in a few conditions. However, no phase separation or crystals were observed.

Since there was no crystallisation lead with the orthologue of Rv3030 from *M. thermoresistibile,* the possibility of obtaining crystals from the *M. smegmatis* orthologue was explored.

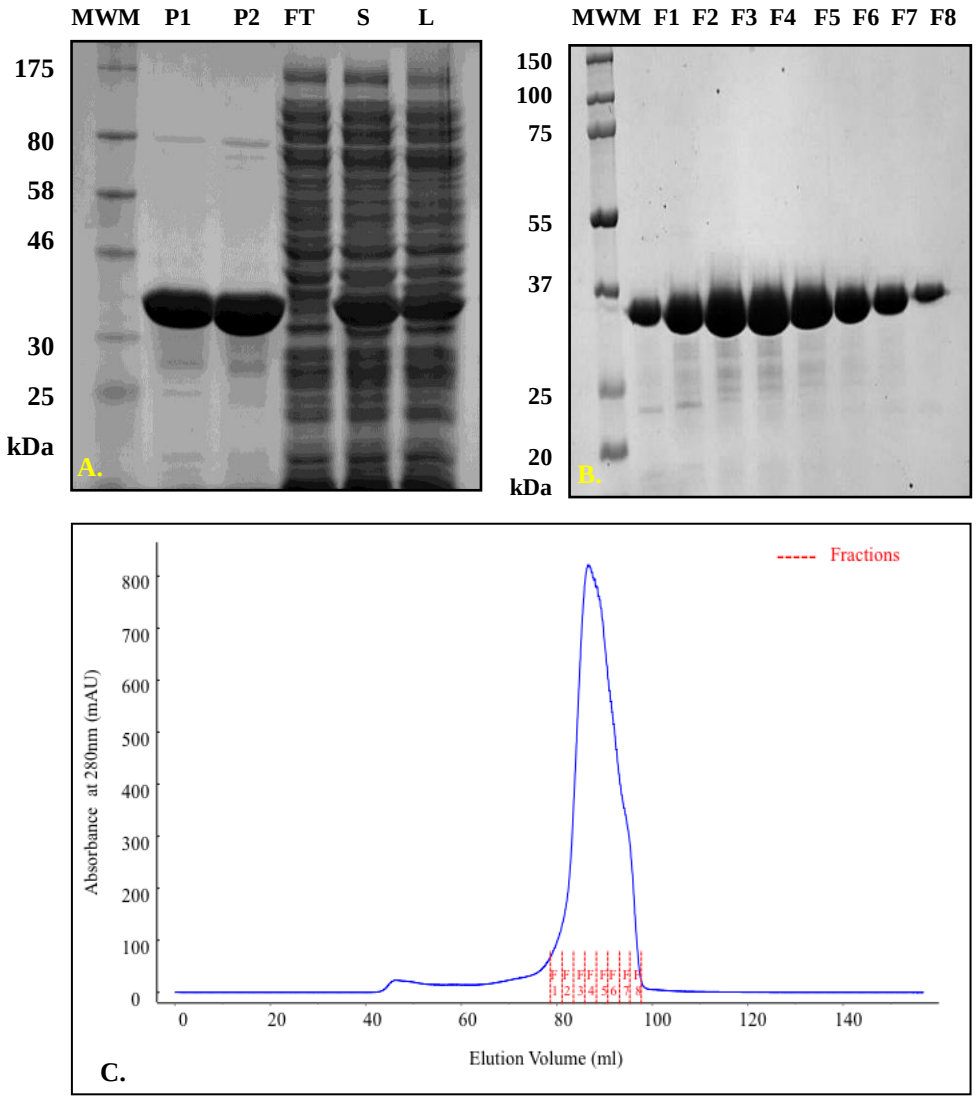


FIGURE 3.4 Expression and purification of C-terminus His_6-tagged orthologue of Rv3030 from *M. thermoresistibile.*

A. Metal affinity purification of His tagged protein on Ni-sepharose column. (MWM: molecular weight marker, P1 and P2: tagged protein, FT: flow through, S: supernatant, L: cellular lysate)

B. Gel filtration of fusion protein on Superdex-200 column. (MWM: molecular weight marker, F1 to F8: protein fractions)

C. The gel filtration profile.

3.5.4 Full-length Orthologue of Rv3030 from *M. smegmatis*

3.5.4.1 Expression and Purification of the Full-Length Orthologue of Rv3030 from *M. smegmatis* (with subsequent tag cleavage)

The 42 kDa recombinant protein with cleavable SUMO tag at its N-terminus was purified by

affinity chromatography and the tag was subsequently cleaved. The trace from the gel filtration run showed that the tag-less protein behaved as a monomer in the solution **(Figure 3.5)**.

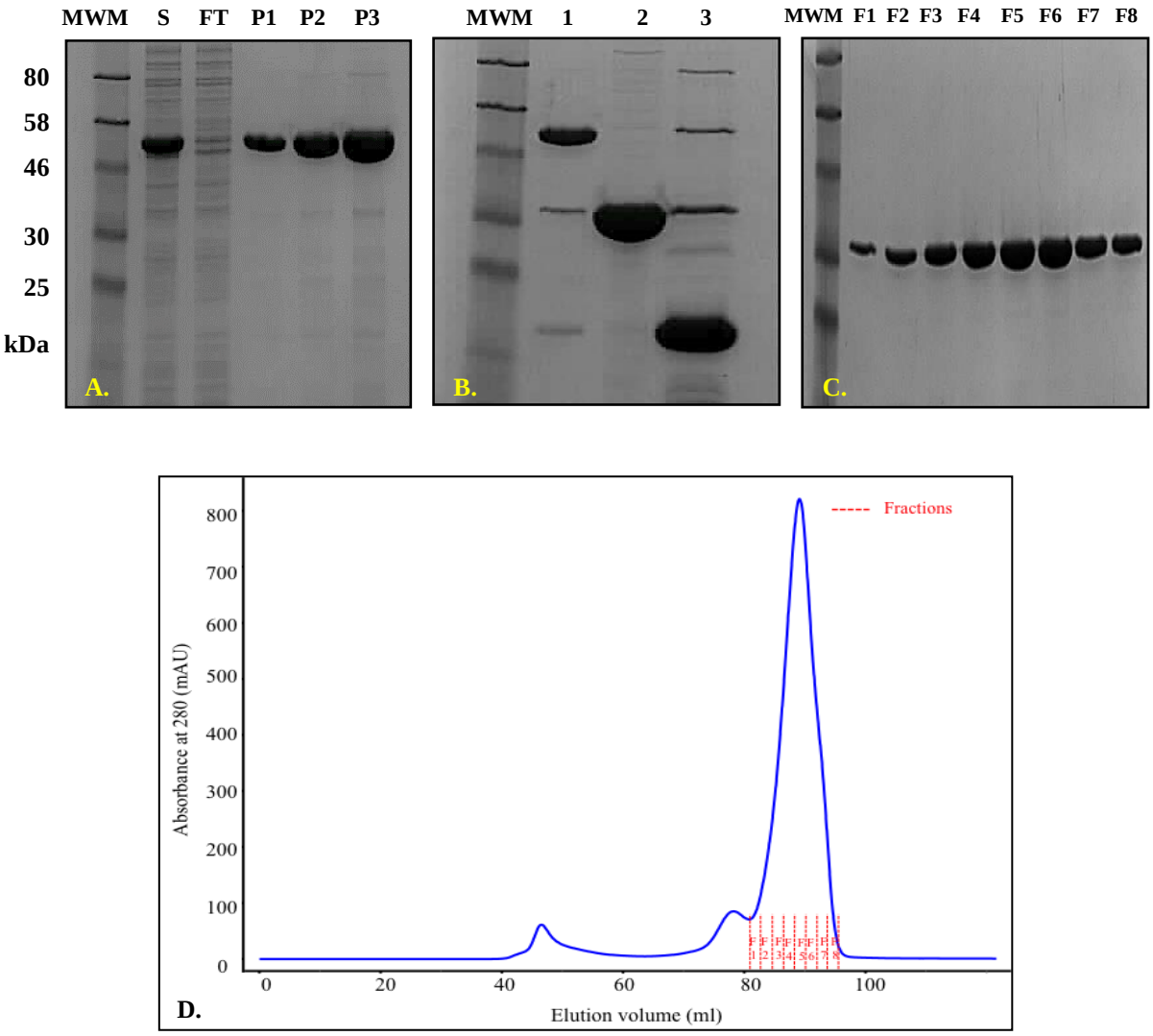


FIGURE 3.5 Expression and purification of full-length orthologue of Rv3030 from *M. smegmatis* (with subsequent tag cleavage).

A. Metal affinity purification of SUMO tagged protein on Ni-sepharose column. A strong band seen at MW higher than that of the protein corresponds to the SUMO-tagged protein. (MW of protein- 28.1 kDa, MW of SUMO tag- 13.8 kDa, total MW of tagged protein- 42 kDa). (MWM: molecular weight marker, S: supernatant, FT: flow through, P1 to P3: tagged protein)

B. Cleavage of SUMO tag by UlpI protease and reverse binding on Ni-sepharose column. (1: tagged protein, 2: tag less protein obtained in the flow through, 3: SUMO tag).

C. Gel filtration of tag-less protein on Superdex-200 column. (MWM: molecular weight marker, F1 to F8: protein fractions).

D. The gel filtration profile.

3.5.4.2 Crystallisation of Full-Length, Untagged Orthologue of Rv3030 from *M. smegmatis*

400 conditions were initially screened with untagged protein at a concentration of 20 mg/ml. Heavy precipitation was observed in the majority of the conditions. However, a phase separation was observed in the condition containing 4 M sodium formate, and the UV trace confirmed the solution with protein was separated **(Figure 3.6)**.

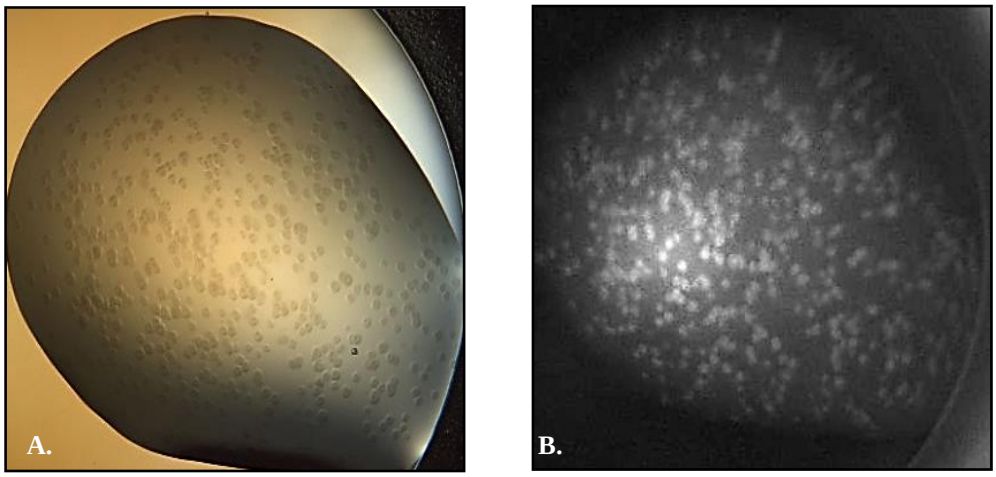


FIGURE 3.6 Phase separation of full-length, untagged orthologue of Rv3030 in *M. smegmatis*.

A. Phase separation in the condition containing 4 M sodium formate.
B. UV image of the same condition confirming separation of protein phase.

The above condition was optimised further to obtain crystals, by creating a pH gradient and concentration gradient of sodium formate, with constant NaCl concentration. Protein at a concentration of 10 mg/ml was used for optimisation. The crystals were formed in the condition containing 4 M sodium formate, 0.2 M NaCl and 0.1 M Tris-HCl pH 8.5. Later, crystals were formed at other pH values but at same sodium formate and NaCl concentrations (4 M and 0.2 M, respectively). Thus, crystal growth was independent of pH, but strictly required 4 M sodium formate **(Figure 3.7)**. However, the crystal condition was not reproducible, and the crystals took very long time to grow (approx. two months). The crystals diffracted to low resolution of 6 Å, which made them not suitable for high-resolution structure determination and fragment-based drug-discovery.

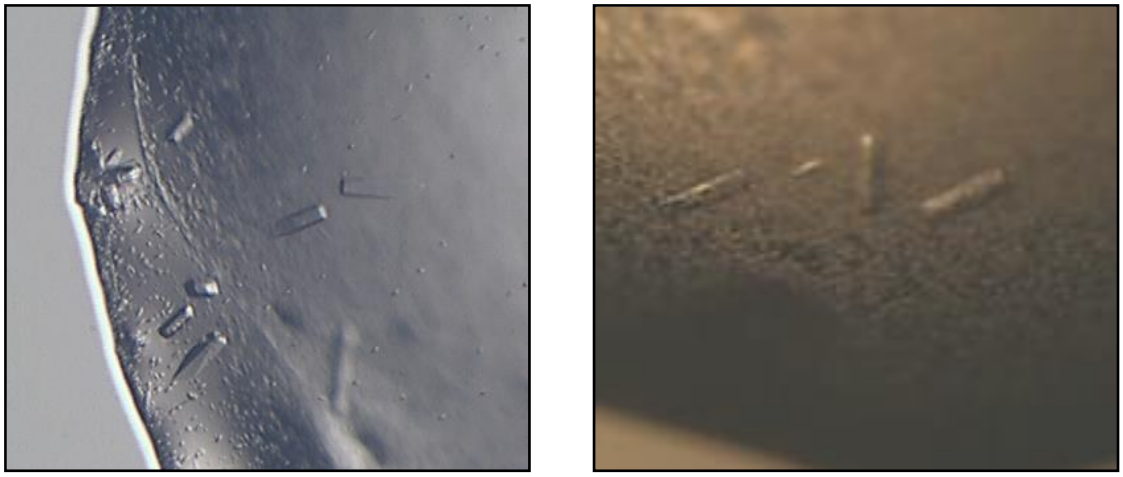


FIGURE 3.7 Crystals of full-length, untagged orthologue of Rv3030 from *M. smegmatis*.

Crystals in condition containing 4 M Sodium formate, 0.2 M NaCl and 0.1 M Tris-HCl at different pH. Crystal formation was invariant of pH. However, the condition was not reproducible and diffracted poorly.

The processed X-ray diffraction dataset showed that crystals belonged to the space group I4 and the unit cell dimensions were:

a=b=114.2 Å; c=268.7 Å

α=β=γ=90°

However, it was not possible to solve the structure by molecular replacement due to lack of a suitable probe. Other phasing techniques, such as using isomorphous replacement by preparing heavy-metal derivatives were not possible due to lack of reproducibility of crystals. Also, seleno-methionine labelling was not feasible because of presence of only one methionine residue in the protein.

An array of conditions was trialled with full-length orthologues in order to improve the resolution of crystals obtained and to identify new crystallisation conditions. These included: setting crystallization experiments with the tagged protein, co-crystallisation of untagged protein with active substrate (SAM) and substrate analogues (SAH, maltohexaose), co-crystallisation of untagged protein with Mg^{2+}, limited proteolysis, additive screening, screening with different salts, streak seeding and matrix seeding. However, all the attempts were unsuccessful.

3.5.4.3 Expression and Purification of the Full-Length Orthologue of Rv3030 from *M. smegmatis* (without subsequent tag cleavage)

High levels of recombinant protein with cleavable SUMO tag at its N-terminus were produced. However, SUMO tag was not cleaved by the Ulp1 protease. The gel filtration profile showed that the SUMO-tagged protein also behaved as a monomer in solution. A single, sharp peak was seen on the trace **(Figure 3.8)**.

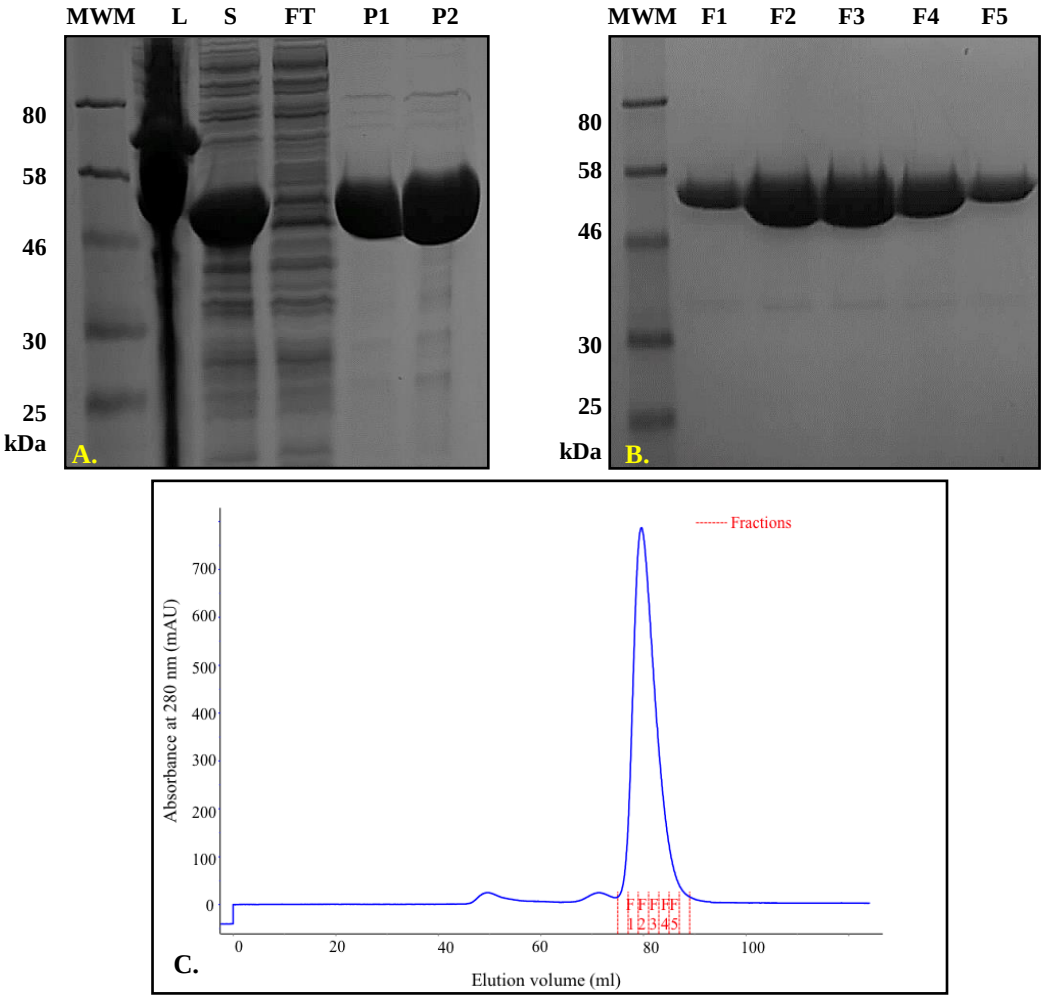


FIGURE 3.8 Expression and purification of full-length, SUMO tagged orthologue of Rv3030 from *M. smegmatis*.

A. Metal affinity purification of SUMO tagged protein on Ni-sepharose column. A strong band can be seen at a MW higher than that of the protein. (MW of protein- 28.1 kDa, MW of SUMO tag- 13.8 kDa, total MW of tagged protein-42 kDa). (MWM: molecular weight marker, P1 to P2: tagged protein, FT: flow through, S: supernatant, L: cellular lysate).

B. Gel filtration of tagged protein on Superdex-200 column. (MWM: molecular weight marker, F1 to F5: protein fractions).

C. The gel filtration profile.

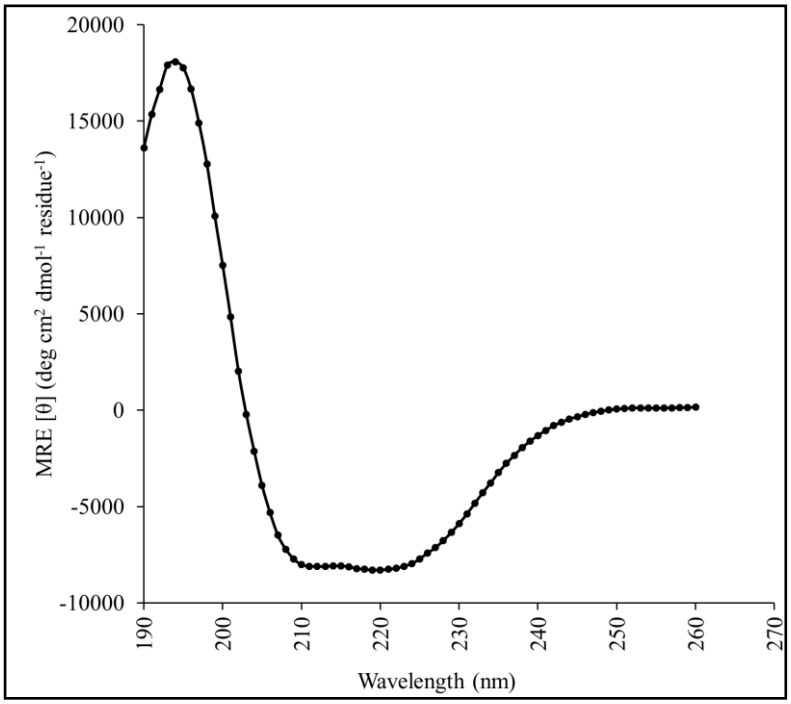
3.5.4.4 Crystallisation Trials with Full-Length, Tagged Orthologue of Rv3030 from *M. smegmatis*

Initially, 400 conditions were screened with the tagged-protein at a concentration of 12 mg/ml. Precipitation was observed in a majority of the conditions but no lead was identified.

3.5.5 Analysis of Secondary Structure by Circular Dichroism

The secondary structure composition of Rv3030 orthologue from *M. smegmatis* was assessed by CD spectroscopy. The UV spectrum showed that full-length protein was folded, and was composed of α-helices and β-strands. The data were deconvoluted using the CDSSTR programme and fitted at normalized RMSD value of 0.014. The protein contains 31% helical content, and a slightly lower β-sheet content of 25% **(Figure 3.9)**.

A.



B.

NRMSD	0.014
α-helix	31%
β-sheet	25%
Turns	20%
Unordered	26%

FIGURE 3.9 The circular dichroism spectrum of full-length orthologue of Rv3030 from *M. smegmatis.*

A. Far UV CD spectra of orthologue of Rv3030 from *M. smegmatis* recorded between 260 and 190 nm and presented in mean residue ellipticity, MRE ([θ]).

B. Deconvolution of the data using the CDSSTR program of DichroWeb.

3.5.6 Determining the Oligomerization State by SV-AUC

The oligomeric state of full-length orthologue of Rv3030 from *M. smegmatis* was determined by SV-AUC. The data were fitted to a continuous c(S) distribution model, which resulted in a sharp, single peak in both the measurements from UV absorption and interference optics **(Figure 3.10)**. The estimated molecular weight of the protein from the absorbance scan was 28.2 kDa; remarkably close to its theoretical monomeric size of 28.1 kDa, and the estimated molecular weight from interference optics was 29.1 kDa, which is within 5% of the theoretical weight. This indicates that the protein exists as monomer in solution, confirming the observation from the gel filtration profile. The frictional ratio calculated from the absorbance and interference experiments was 1.27 and 1.32, respectively. This indicates that the protein adopts a globular conformation.

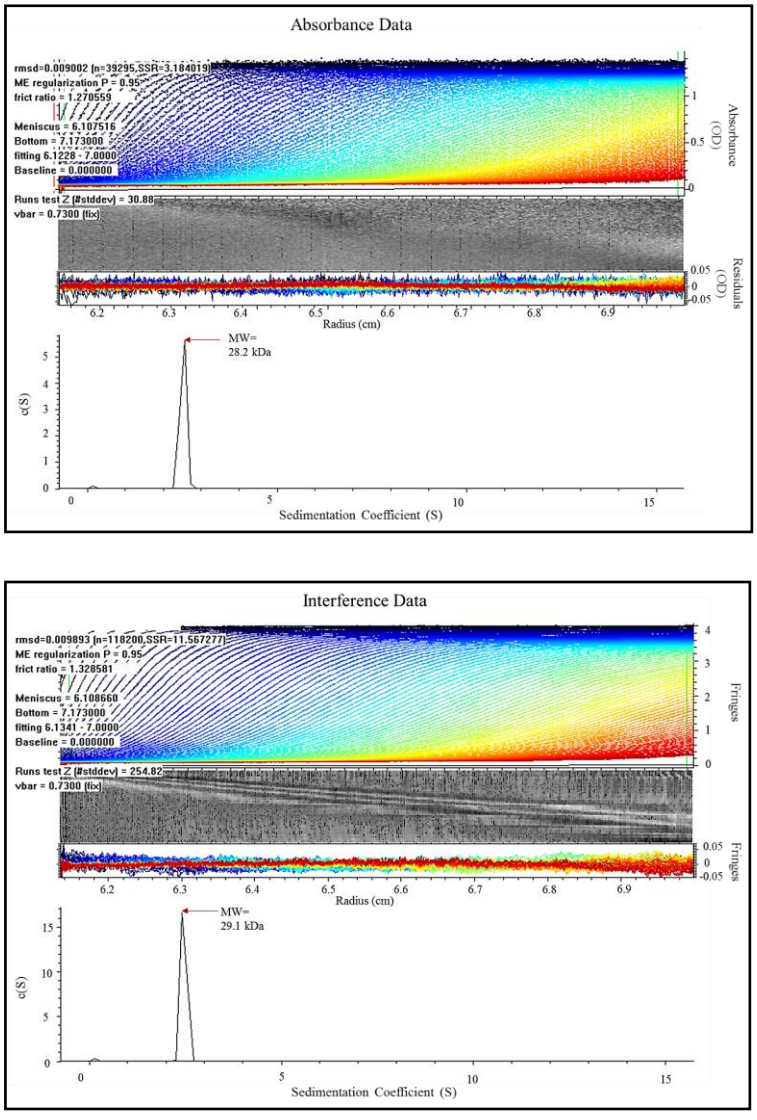


FIGURE 3.10 The SV-AUC analysis of full-length orthologue of Rv3030 from *M. smegmatis.*

3.5.7 Characterisation of Ligand Binding to the Full-Length Orthologue of Rv3030 from *M. smegmatis*

The binding of SAM, the methylgroup donor, to this 6-O methytransferase was observed by both ligand-NMR and ITC. A decrease in intensity of ^{1}H signal was observed when SAM bound to the protein **(Figure 3.11 A)**. The thermodynamics of binding of the co-factor was determined, and by fitting the ITC data, K_d was observed to be 41.5 µM ± 3.7 µM. The complexation of SAM with this methyltransferase was exothermic (negative peaks in the ITC output), with 1:1 stoichiometry **(Figure 3.11 B)**. The association was characterised by a favourable enthalpy change and a negative entropy change (values of thermodynamic properties are shown in an inset in the figure 3.11)

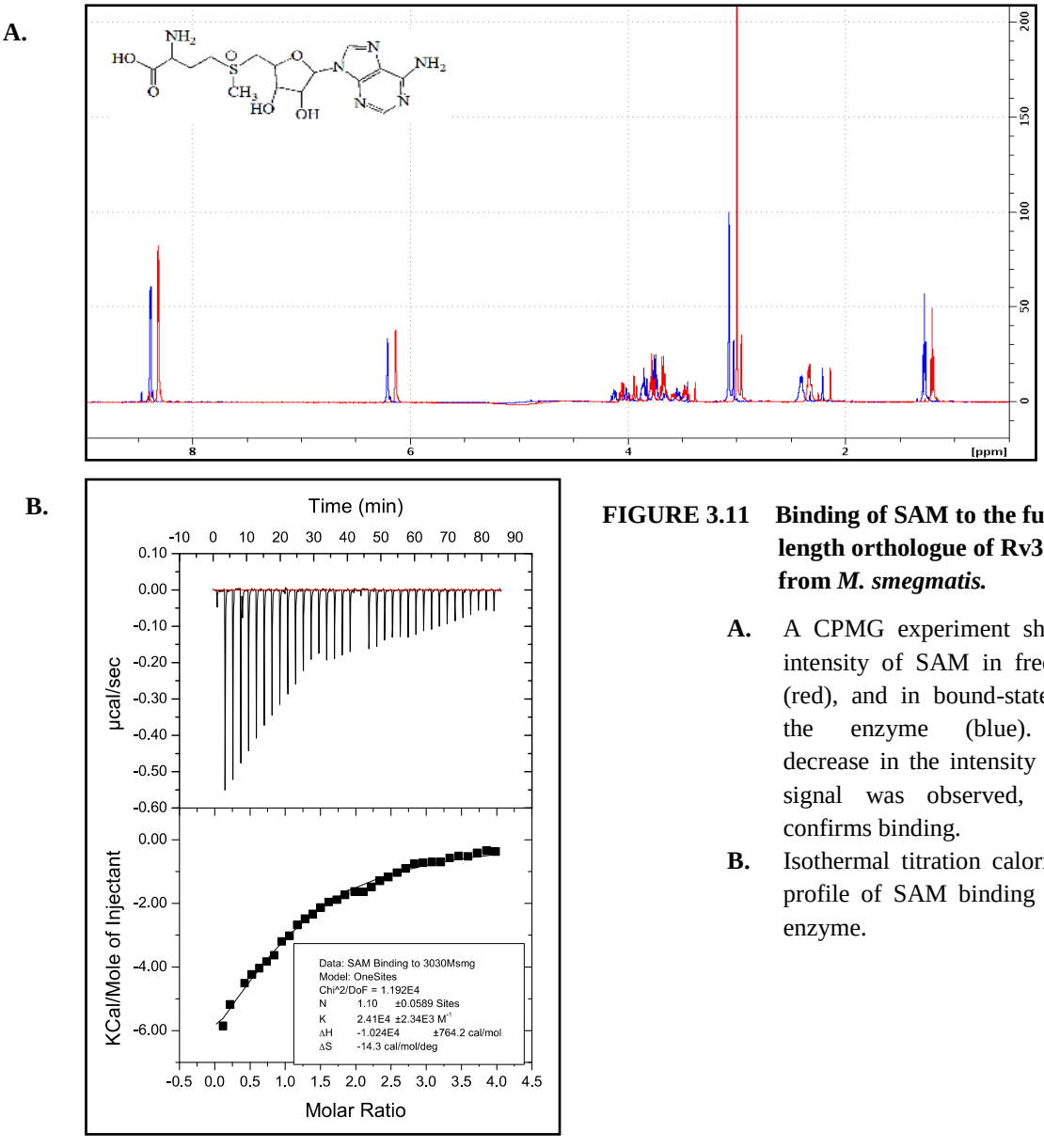


FIGURE 3.11 Binding of SAM to the full-length orthologue of Rv3030 from *M. smegmatis.*

A. A CPMG experiment showing intensity of SAM in free-state (red), and in bound-state with the enzyme (blue). The decrease in the intensity of the signal was observed, which confirms binding.

B. Isothermal titration calorimetry profile of SAM binding to the enzyme.

3.5.8 N-terminus Truncations of Orthologue of Rv3030 from *M. smegmatis*

The homology model of the orthologous *M. smegmatis* protein indicated that some residues from the N-terminus of the protein do not align to any secondary structure element. This was concordant with the secondary structure prediction results obtained from PSIPred **(Figure 3.2)**. Therefore, these residues from the unstructured loop at the N-terminus were truncated, and this approach was followed so as to readily crystallise and obtain the structure of rest of the protein. Six different constructs which had 4, 8, 14, 15, 17, and 24 residues truncated respectively, were screened.

3.5.8.1 Expression and Purification of the Truncated Orthologues of Rv3030 from *M. smegmatis*

All the constructs with cleavable N-terminus SUMO tag were successfully expressed and purified from the soluble fraction with high purity. The fusion tag was subsequently cleaved and final polishing was done by gel filtration on Superdex-200 column. The gel-filtration profile demonstrated that all the proteins behaved as monomer in solution **(Figure 3.12 and Figure 3.13)**.

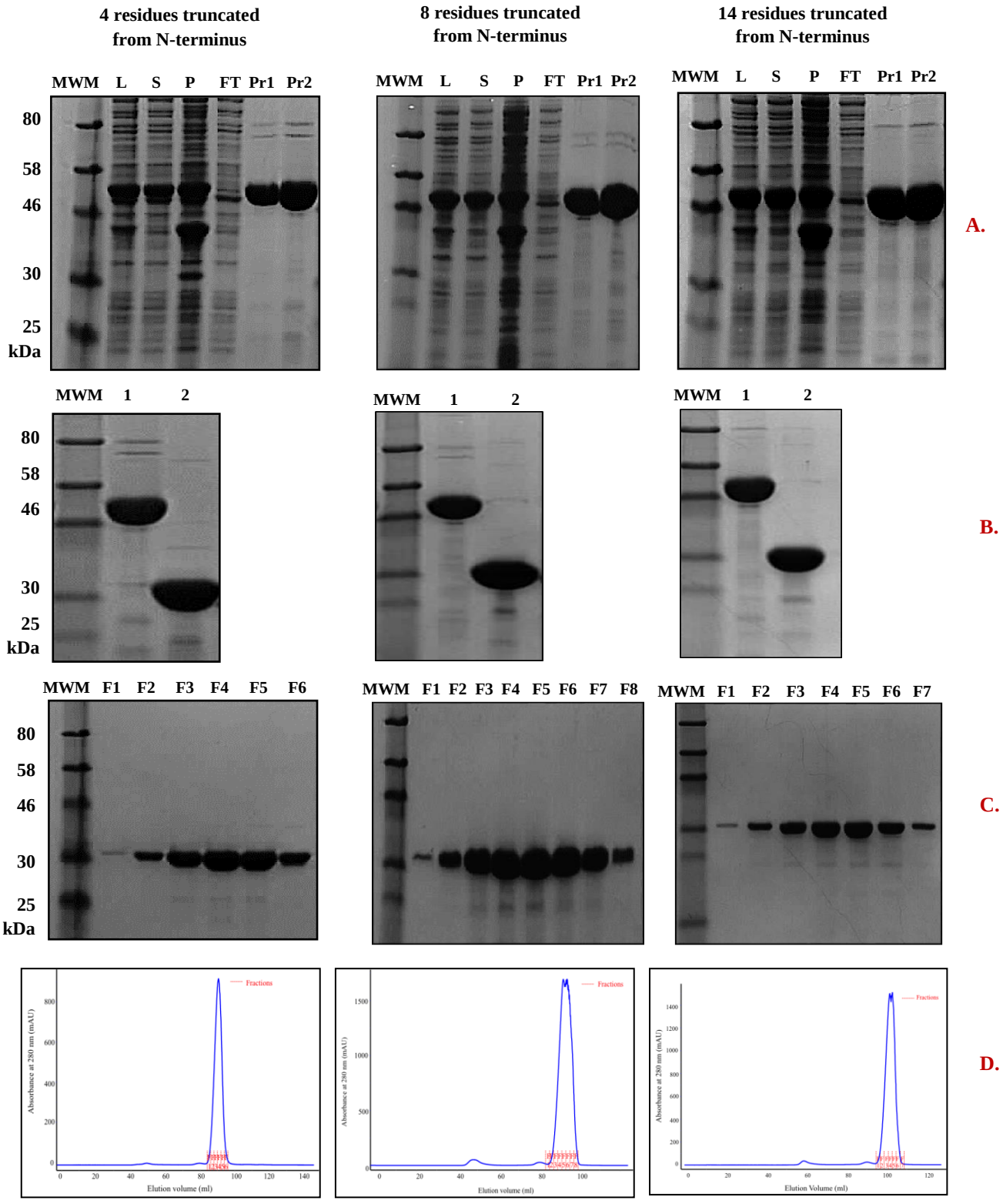


FIGURE 3.12 Expression and purification of truncated orthologues of Rv3030 from *M. smegmatis*.

A. Metal affinity purification of SUMO tagged proteins on Ni-sepharose column. Strong bands seen at MW higher than that of the proteins. (MWM: molecular weight marker, L: cellular lysate, S: supernatant, FT: flow through, Pr1 to Pr2: tagged proteins).

B. Cleavage of SUMO tag by UlpI protease and reverse binding on Ni-sepharose column. (1: tagged proteins, 2: tag less proteins obtained in the flow through).

C. Gel filtration of tag-less proteins on Superdex-200 column. (MWM: molecular weight marker, F1 to F8: protein fractions).

D. Gel filtration profiles.

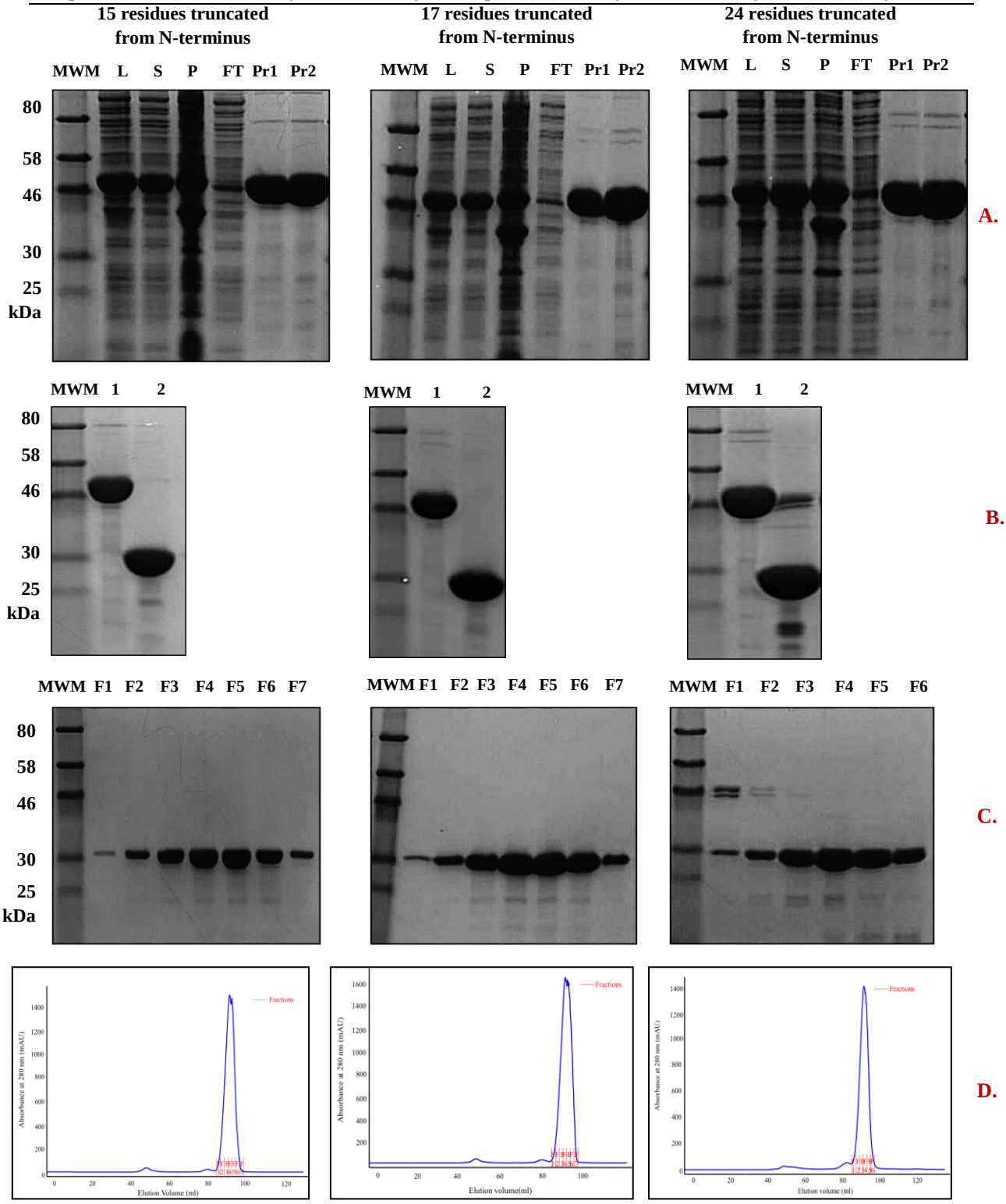


FIGURE 3.13 Expression and purification of truncated orthologues of Rv3030 from *M. smegmatis.*

A. Metal affinity purification of SUMO tagged proteins on Ni-sepharose column. Strong bands seen at MW higher than that of the proteins. (MWM: molecular weight marker, L: cellular lysate, S: supernatant, FT: flow through, Pr1 to Pr2: tagged proteins).

B. Cleavage of SUMO tag by UlpI protease and reverse binding on Ni-sepharose column. (1: tagged proteins, 2: tag less proteins obtained in the flow through).

C. Gel filtration of tag-less proteins on Superdex-200 column. (MWM: molecular weight marker, F1 to F7: protein fractions).

D. Gel filtration profiles.

3.5.8.2 Crystallisation Trials of Truncated Orthologues of Rv3030 from *M. smegmatis*

Each of the truncated proteins was screened against approximately 600 crystallisation conditions, but no phase separation or crystals were observed. However, for one of the truncated proteins, which was truncated by 17 residues at its N-terminus, microcrystals were formed but in only one condition containing 0.1 M MES pH 6.5, 1.8 M $AmSO_4$, 0.01 M cobalt (II) chloride **(Figure 3.14 A and B)**. Further optimisation of this condition by additive screening led to identification of a phase separation condition (0.1 M MES pH 6.5, 1.8 M AmSO4, 0.01 M cobalt (II) chloride, 0.5% PVP-K15), which was confirmed to be of protein by UV imaging **(Figure 3.14 C and D)**. In spite of considerable optimisations carried out, no crystals were formed for this truncated construct.

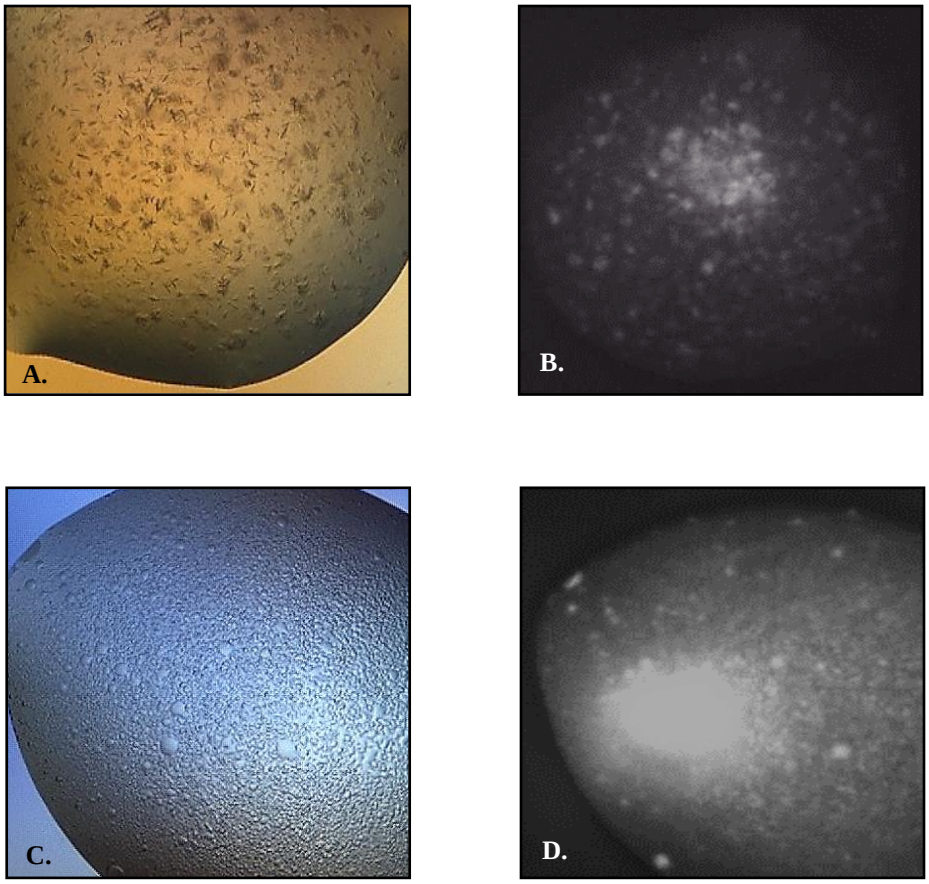


FIGURE 3.14 Microcrystals and phase separation of truncated (17 residues from N-terminus) orthologue of Rv3030 from *M. smegmatis.*

- **A.** Microcrystals from initial screening.
- **B.** UV image of the same condition confirming microcrystals.
- **C.** Phase separation after optimization of microcrystalline condition.
- **D.** UV image of the same condition confirming separation of protein phase.

3.6 Discussion

Rv3030, a SAM-dependent methyltransferase, carries out methylation at the 6-O position of glucoses, which are vital for synthesis of mature MGLP chains. Sequence alignment of Rv3030 and its orthologues demonstrated that the enzyme was conserved among the species of mycobacteria, thus suggesting that either of the orthologues from *M. smegmatis* or *M. thermoresistibile* could be used as a surrogate for structural and drug-discovery studies. The three-dimensional structure of the full-length orthologue of Rv3030 from *M. smegmatis* was modelled. This suggested that the protein structure comprises a single domain, with an unstructured N-terminus, consistent with the results of *in silico* prediction of a disordered region of the protein

The results of protein expression and purification show that good quantities of full-length soluble proteins, purified to homogeneity, were obtained from both *M. smegmatis* and *M. thermoresistibile*. The crystallisation of full-length proteins was carried out. Notwithstanding, crystallisation proved challenging, and there was no success in obtaining well diffracting crystals suitable for structure determination. In spite of testing and optimising a number of conditions, the best crystals produced were of full-length *M. smegmatis* orthologue. However, these were formed in only one condition, took long to form and diffracted only to a very low resolution (6Å). Processing of the low-resolution dataset of the full-length orthologue of Rv3030 from *M. smegmatis* showed that the crystals belonged to the I4 space group. Attempts to solve the structure were not successful due to several factors. Firstly, the absence of the structure of a close orthologue to use as a probe meant that molecular replacement could not be used. Secondly, preparation of heavy metal derivatives for isomorphous replacement was not feasible, as the crystals could not be reproduced. Thirdly, the presence of just one methionine residue in the protein made it unsuitable for seleno-methionine labelling for single anomalous diffraction (SAD) phasing.

The biochemical characterisation had not been possible due to lack of availability of active substrate of the enzyme. Further biophysical studies were carried out which revealed that the circular dichroism spectrum of the full-length orthologue of Rv3030 from *M. smegmatis* displayed attributes of a folded protein, and SV-AUC data showed that the protein exists as a monomer and is globular. These results affirmed that difficulties in crystallisation of the full-length protein were not because of its incorrect folding. The binding of the co-factor, SAM,

to the enzyme, positively supported that the enzyme was folded, as the binding affinity observed was biologically relevant.

Based on the observations from the *in silico* model of the structure of the Rv3030 orthologue from *M. smegmatis*, a rationale was derived for manipulation of the gene and truncation of the disordered region at the N-terminus of the protein, in order to make it amenable for crystallisation. Several constructs, with varying degrees of truncation of the unstructured region were designed and screened. All the truncated proteins were successfully expressed and purified from the soluble fraction, with high purity. However, attempts to crystallise these did not give rise to well-formed, high-resolution crystals.

Although the results of biophysical assays presented in this chapter show that the protein is folded, crystallisation of full-length and truncated proteins was very demanding. Henceforth, alternative approaches to study the structure of the protein were explored. These included nuclear magnetic resonance (NMR) and small-angle X-ray scattering (SAXS), described in the next chapter.

CHAPTER-4

Probing the Structural Features of Orthologue of Rv3030 from *M. smegmatis* by SAXS and NMR Spectroscopy

4.1 Summary

At the outset of this project, crystallisation was pursued to determine the structure of the enzyme. However, as presented in chapter-3, crystallisation was not tractable. Thereby, other possibilities were sought to understand the structural features of Rv3030 using its orthologue in *M. smegmatis* as the representative model, and insights were gained from small-angle X-ray scattering and NMR spectroscopy. The results demonstrated that the enzyme was folded and globular, confirming the observations from the biophysical investigation. Further, NMR-based protein dynamic study revealed that the N-terminus is intrinsically disordered. The secondary structure characterisation of protein by NMR showed that the protein had the β-sheet topology broadly classified as a Rossmann fold. These data provide the first evidence about the structural features of this 6-O methyltransferase involved MGLP biosynthesis.

4.2 Introduction

Given the central importance of Rv3030 in the survival of *M. tb*, it is essential to understand the structural properties of the methyltransferase, for further elucidation of the catalytic mechanism and development of specific inhibitors against this target. Although well-formed crystals of full-length orthologue of Rv3030 were obtained, they were not of suitable diffraction quality for structure determination. Therefore, a SAXS and NMR-based study to determine the structure of the protein was undertaken.

4.3 Basic Principles of Small-Angle X-ray Scattering and Nuclear Magnetic Resonance

4.3.1 Small-Angle X-Ray Scattering

Small-angle X-ray scattering (SAXS) is a technique used to study structure and measure biophysical parameters of macromolecules in solution. It provides information about the size, shape, molecular assembly and conformation of the proteins, other biological molecules and complexes. It is relatively a fast and straightforward technique, with no limitation to the protein size. The technique also allows quantitative analyses of flexible and disordered regions of proteins. Further understanding on the technique can be gained by referring to the texts: Graewert and Svergun, 2013; Kikhney and Svergun, 2015; Mertens and Svergun, 2010; Petoukhov and Svergun, 2013; Svergun and Koch, 2003.

The protein solution in quartz capillary is placed in a monochromatic X-ray beam. The intensity of scattered X-rays is measured and recorded by an x-ray detector **(Figure 4.1)**.

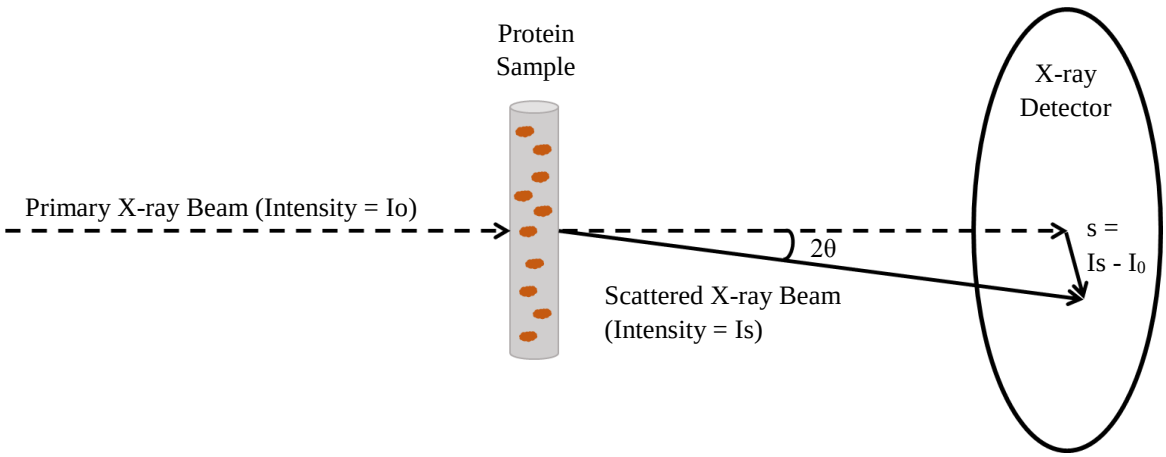


FIGURE 4.1 Basic principles of a SAXS experiment.
(s: momentum transfer) (Figure adapted from Kikhney and Svergun, 2015)

Pure solvent is used as the reference, and the scattering pattern collected for it is subtracted from the sample solution. The resulting pattern gives the overall size and shape of the protein.

The protein molecules in the solution are randomly oriented, and thus, the scattering pattern is isotropic, and the intensities and amplitudes of scattered beam are spherically averaged.
The momentum transfer (*s*) between the primary and scattered X-ray beam is defined as

$$s = \frac{4\pi sin\theta}{\lambda}$$

where, λ – wavelength
2θ – scattering angle

The atomic scattering amplitude, which is the Fourier transformation of the electron density, is represented by the following equation:

$$A(\boldsymbol{s}) = \int (\rho(r) - \rho_s) e^{(isr)} d\boldsymbol{r}$$

where, $\rho(r)$ – electron density of particle
ρ_s – electron density of solvent
r – vector between atoms
s – |**s**|

SAXS provides valuable information about the characteristics of the protein, including molecular weight, maximum dimension (D_{max}), and radius of gyration (R_g). R_g provides the overall size of the molecule, and for compactly shaped proteins, its value is small.
By using Guinier approximation, the intensity is defined as:

$$I(s) = I(0) \exp(-\frac{s^2 R_g^2}{3})$$

The Guinier plot, where the scattering intensity *I(s)* is plotted against s^2, gives information about the aggregation of particles in the solution. Lack of linearity is a sign of aggregation or repulsive inter-particle interactions.

The distances between all feasible pairs of atom is represented in form of a histogram, known as a distance distribution function, p(r). For a globular protein, the curve is perfectly bell-shaped, whereas there is an extended tail for an unfolded protein.

The following equation helps in calculating R_g from the p(r) function more precisely:

$$R_g^2 = \frac{\int_0^{D_{max}} r^2 p(r) dr}{2 \int_0^{D_{max}} p(r) dr}$$

where, D_{max} – maximum dimension

SAXS data also provide qualitative identification of the flexible and disordered regions of the protein. The Kratky plot ($s^2I(s)$ vs. s) is used to identify the folding state and flexible regions. The plot is bell-shaped for a folded protein, and in case of unfolded proteins, it exhibits as a plateau for a range followed by a constant increase.

Some beamlines at the synchrotrons allow size-exclusion chromatography to be performed just before collecting the SAXS data, which removes unwanted aggregation. This is particularly useful to remove noise signals produced because of disproportionate scattering of the beam, and consequently affecting the data analysis.

4.4.2 Nuclear Magnetic Resonance

This section summarises the basic theory of nuclear magnetic resonance, but does not provide a substantial coverage on the subject matter, which can be found in a number of texts (Kovermann *et al*., 2016; Kwan *et al*., 2011; Tengel, 2007; Williamson, 2013; Wüthrich, 1986, 2001; Zuiderweg, 2002).

Nuclear magnetic resonance (NMR), discovered by Purcell and Bloch in 1946, is a powerful technique to study structure and dynamics of various biomolecules. NMR is complementary to X-ray crystallography and allows characterisation of protein dynamics in solution state. This is particularly advantageous as the dynamics can be quantified over a large time scale and under equilibrium conditions, using multiple probes simultaneously. Solution-state NMR offers sample conditions that can be equated to the physiological conditions in the cell. The size is, however, the limitation for NMR study, and typically proteins < 300 residues are suitable for the structure determination in the absence of costly deuterium-labelling methods.

4.3.2.1 Magnetisation and Resonance

NMR is dependent on interaction of nuclear spin of nuclei of atoms with the magnetic field. The nuclear spin of nuclei having even number of neutrons and protons is zero, whereas it is non-zero for nuclei with odd number of neutrons or protons. Atoms like ^{1}H, ^{13}C, ^{15}N, ^{19}F and ^{31}P, possess nuclear spin of ½ and are exploited for structure determination. The nuclear spin is defined by magnetic quantum number, *I*, which can have values of + ½ and - ½. These values correspond to orientation of nuclei either parallel or anti-parallel to the magnetic field, and have different energy states. The parallel orientation has lower energy and is known as the α-state; the anti-parallel orientation, on the other hand, has higher energy and is designated as the β-state.

Resonance occurs when a system at a particular energy state is excited with a specific frequency, known as the resonant frequency, and this results in transition between the two levels having different energies.

The difference in energy between two levels is defined by the following equation:

$$\Delta E = hv$$

where, ΔE – energy difference
h – Planck's constant (6.626×10^{-34} J.s)
v – frequency

The frequency, v, of transition is represented by Larmor equation, and is known as Larmor frequency.

$$v = \frac{\gamma B}{2\pi}$$

where, γ– gyromagnetic ratio
B – magnetic field

The bulk magnetisation (**M**) is defined as a vector sum of the individual nuclear spins in the magnetic field.

The relationship between magnetisation with the magnetic field is given by:

$$\frac{d\boldsymbol{M}}{dt} = -\gamma \boldsymbol{M} \times \mathbf{B} = -\boldsymbol{M} \times \omega$$

where, M – bulk magnetisation

ω – angular Larmor frequency (ω= 2πν)

At the equilibrium state, the bulk magnetisation is parallel to the static external magnetic field (B_0, by convention along the z-axis) resulting in the lowest energy state being marginally favoured. Classically, if the bulk magnetisation is re-oriented away from this position, it will precess around the z-axis at its Larmor frequency. Quantum mechanically, the nuclei transit between different energy states when a radio frequency (RF) pulse of energy equal to ΔE is applied (i.e. at the resonance frequency). Eventually, the bulk magnetisation gets out of phase, due to the effect of T_2 relaxation, and returns back to the equilibrium, due to T_1 relaxation. The decay due to T_2 relaxation is measured as a fluctuating decline of the amplitude of the magnetic waves as a function of time, known as free induction decay (FID). Fourier transformation of the FID converts it from a function of time to a function of frequency, and the data is represented as an NMR spectrum.

4.3.2.2 Chemical Shift

The surrounding electrons can marginally alter the magnetic field experienced by the nuclei of the same atoms. It is because when external magnetic field is applied, the electrons also rotate in their atomic orbitals and produce localized magnetic fields aligned opposite to the direction of the magnetic field. As a result, the strength of the local magnetic field and the resonant frequency of the nuclei are reduced. This shielding effect of electrons is influenced by the chemical environment and is proportional to the magnetic field.

Chemical shift is defined as the resonant frequency of a sample nucleus relative to that of a reference in the magnetic field, and is expressed in ppm, by the following equation:

$$\delta = \frac{(v - v_{ref})}{v_{ref}}$$

where, δ – chemical shift

v – resonant frequency of sample nucleus

v_{ref} – resonant frequency of reference

Generally, the reference used is the resonant frequency of ^{1}H nuclei of tetramethylsilane (δ_{TMS}= 0 ppm). Proteins have characteristic proton chemical shifts, which helps in the identification of various structural groups.

4.3.2.3 J- Coupling

J-coupling is also known as scalar coupling or indirect spin-spin coupling. In a system of covalently bound nuclei, the local magnetic field experienced by a particular spin is not only dependent on the surrounding electrons, as in the chemical shift, but is also dependent on the spin states of the neighbouring nuclei, arbitrated by electrons through the chemical bonds. When an adjacent nuclear spin is parallel to the magnetic field, the local magnetic field is increased at the local nucleus. Conversely, it is decreased if the adjacent spin is in an anti-parallel orientation. The J-coupling is independent of the external magnetic field and is mutual, i.e. the same for coupled nuclei.

This phenomenon splits the NMR signal into multiple lines, depending upon the number of surrounding nuclei, which in protein NMR are mostly not helpful. Therefore, the heteronuclear spin couplings are decoupled or selectively reduced. However, J-coupling provides useful information about dihedral angles and bond lengths that is used for model building. The magnitude of coupling, known as the coupling constant J, indicates the proximity of the coupled nuclei. As the number of bonds increases, the magnitude generally decreases and vice-versa.

4.3.2.4 Heteronuclear NOE

The backbone ^{1}H-^{15}N heteronuclear NOE provides information about the motion of the individual N-H bond vectors. Those that undergo motion faster than the overall tumbling of the molecules (i.e. in the pico-to-nanosecond time scale) show a decreased NOE intensity relative to the average observed for the majority of the residues. Thus, for instance, decreased values are usually found at both N- and C-terminal ends of the protein (< 0.6), compared with ordered regions (0.6-0.9).

4.3.2.5 Multi-Dimensional NMR and Resonance Assignment

Multi-dimensional NMR and isotope labelling has enabled determination of macromolecular structures of high molecular masses 25 kDa or higher. This is primarily because the resolution of the multidimensional spectra is better than the corresponding one-dimensional spectra. In the 1-D spectrum, the ^{1}H frequencies are overlapped which makes the assignment of resonances unfeasible, whereas in multidimensional spectra, the frequencies are recorded in multiple dimensions, thereby the peaks are resolved and assignment of individual peaks is possible. Multi-dimensional NMR relies on the transfer of magnetisation between nuclei, which allows the relationships between different nuclei to be mapped out.

The transfer of magnetisation occurs through different methods outlined below:

a) Through-bond correlation methods, wherein the magnetisation is transferred through J-coupling between bonded nuclei, either of the same type (homonuclear) or of different types (heteronuclear). Homonuclear methods include COSY and TOCSY, whereas HSQC is a heteronuclear method.

b) Through-space correlation methods, which transfer magnetisation between nuclei with physical proximity, irrespective of any chemical bond between them. These methods are based upon the nuclear Overhauser effect (NOE).

2D ^{1}H-^{15}N HSQC Spectrum

^{15}N-HSQC (heteronuclear single quantum coherence spectroscopy) is a standard 2D experiment, which serves as a structural fingerprint and forms an important basis for the heteronuclear analysis of protein. It shows correlations between nitrogen and proton nuclei of the residues, and the magnetisation is transferred through the J-coupling and often restricted to the value of a single bond by setting the transfer time relatively short. One peak for every amide spin pair is seen; amine groups are rarely observed due to rapid exchange of the proteins with the solvent. An exception to this is proline, which lacks an amide proton. Predominantly, backbone amide groups are detected, however, peaks are also produced for Trp side-chain Nε-Hε, Asn side-chain Nδ-Hδ2 and Gln side-chain Nε-Hε2. Dependent on their exchange properties, peaks for Arg Nε-Hε may also appear, but the chemical shift for Nε lies outside the usual spectral range, and so the Nε may appear at a frequency ±1 spectral width from the true frequency. This is known as folding or aliasing of the peaks. The same effect is observed for Arg Nη-Hη and Lys Nζ-Hζ, when working at low pH.

The HSQC spectrum provides useful information about the protein folding. When the protein is folded, the peaks are well scattered and individual peaks may be resolved. On the contrary, disordered proteins produce a spectrum that has a cluster of highly overlapped peaks in the central region, which makes the resonance assignment difficult. ^{15}N-HSQC is a major factor in analysing the suitability of a protein for NMR structure determination. It is also applicable in identification of hot-spots of interactions and, thus, mapping the binding surface in protein-drug, protein-ligand and protein-protein interaction, by recording the changes in the chemical shifts of the residues in apo-state with those in ligand-bound state. This is known as chemical shift perturbation mapping.

Triple Resonance Experiments in Backbone and Sidechain Assignment

These experiments are based on bonded connections between ^{1}H, ^{15}N and ^{13}C atoms present in the residues. The protein residues contain the basic H-N-Cα-C' unit, repeated in the polypeptide chain, which is exploited for the resonance assignment. Several spectra are recorded, which give both intra- and inter-residue connections, and rely on transfer of magnetization by J-coupling between N-H, C-C, C-H and C-N bonds. The chemical shift is spread across the three dimensions involving both backbone and side-chain H, N and C nuclei.

The information about backbone Cα for the i^{th} and $(i-1)^{th}$ residue is revealed by an HNCA spectra, and an HN(CO)CA spectra informs about the $(i-1)^{th}$ residue. In the HNCA experiments, the magnetisation is transferred from amide proton to amide nitrogen and then to ^{13}Cα atoms of both the i^{th} and $(i-1)^{th}$ residues via N-Cα J-coupling. In the HN(CO)CA experiment, the magnetisation transfers through the carbonyl carbon atom, and thus, only the peak of ^{13}Cα atom of $(i-1)^{th}$ residue can be observed **(Figure 4.2 A and B)**. The magnetisation in both these experiments then returns back to the amide proton, and these are known as 'out and back' experiments.

Similarly, a pair of 'out and back' experiments, HNCACB and HN(CO)CACB, are recorded for Cβ resonances. The HNCACB experiment records ^{13}Cα and ^{13}Cβ peaks for both the i^{th} and $(i-1)^{th}$ residues, whereas the HN(CO)CACB experiment detects only the peaks for the $(i-1)^{th}$ residue. In the HNCACB experiments, the magnetisation is transferred from amide proton to amide nitrogen and then to ^{13}Cα and ^{13}Cβ atoms of both the i^{th} and $(i-1)^{th}$ residues. In the HN(CO)CACB experiment, the magnetisation transfers through the carbonyl carbon atom, and thus, only peaks of ^{13}Cα and ^{13}Cβ atoms of $(i-1)^{th}$ residue can be observed **(Figure 4.2 C and D)**.

An HC(CCO)NH experiment can also be used for backbone assignment. In this experiment, the magnetisation is transferred from the protons of the sidechain carbons to their ^{13}C atom, via isotropic mixing between all the carbons. From here, the magnetisation is transferred, through the carbonyl carbon, to the amide nitrogen and amide proton of the $(i+1)^{th}$ residue **(Figure 4.2 E)**. The backbone carbonyl is detected through the HNCO spectrum, in which the

magnetisation is transferred from amide proton to amide nitrogen and then to the carbonyl carbon atom **(Figure 4.2 F)**.

An interplay of these spectra leads to the sequential identification of residues and backbone resonance assignment. However, the sequential assignment is interrupted by the presence of proline in the protein sequence, as prolines do not have an amide proton.

It is also required to assign the side-chains to produce complete set of assignments, essential for generating geometric constraints and structural calculations. Typical experiments performed for side-chain assignments are HCCH-COSY and HCCH-TOCSY, wherein magnetisation transfer occurs via J-coupling and isotropic mixing, respectively.

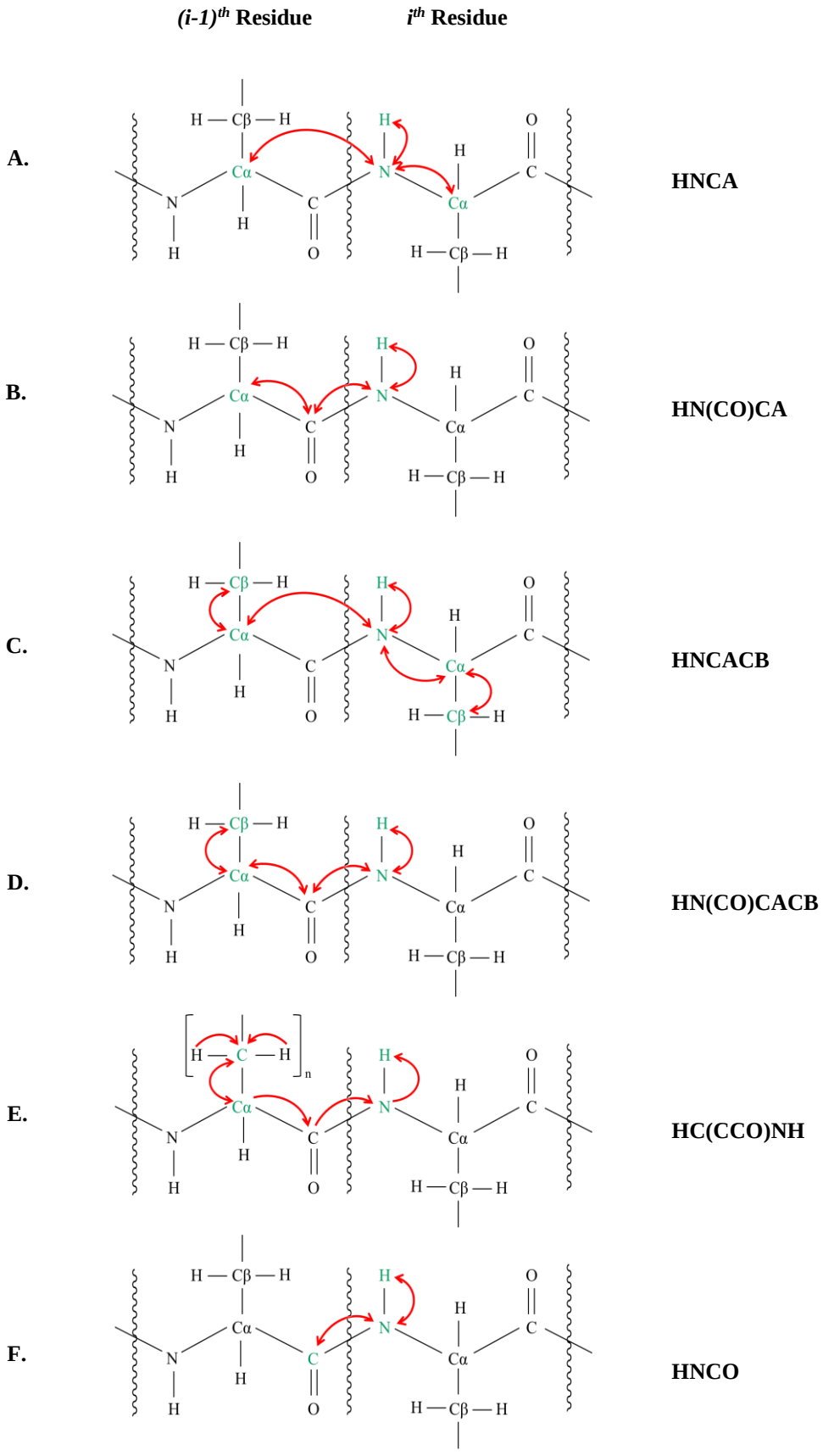


FIGURE 4.2 Triple resonance experiments for backbone resonance assignment.
The red arrows represent the path of magnetisation transfer. The atoms for which the resonance peaks are observed in the spectrum are coloured green.

4.4 Methodology

4.4.1 Small Angle X-ray Scattering of the Full-Length Orthologue of Rv3030 in *M. smegmatis*

The purified protein sample at a concentration of 10 mg/ml in 100 mM NaCl, 20 mM Tris-HCl pH 7.5 was used to collect data. The data were collected at the SWING beamline (Soleil Synchrotron, France). The sample was thawed and centrifuged (20,000 x g, 5 min, room temperature). The sample was placed in a thermo-regulated quartz cell at 15˚C, and loaded onto a Superdex-200 column (GE Healthcare, increase 3.2/300), equilibrated with 100mM NaCl and 20mM Tris-HCl pH 7.5. The sample was eluted directly into the detection chamber, and illuminated with monochromatic X-ray beam of a wavelength of 1.022 Å. The sample was placed 1.79 m from the detector, and a total of 250 measurements were collected. Scattering measurements were collected for the protein buffer (100 mM NaCl, 20 mM Tris-HCl pH 7.5) also, as the reference.

The FOXTROT program was used to radially average the buffer-frames and subtract the averaged buffer-frame from each sample-frame. The buffer corrected sample-frames were averaged, and a final scattering curve was generated for analysis. The data were analysed using the ATSAS package (Petoukhov *et al*., 2012). The R_g value was determined from the slope of Guinier plot in PRIMUS wizard of the package (Konarev *et al*., 2003). The value of D_{max} was calculated from the p(r) curve, by allowing it to reach back to zero. The *ab initio* model of the protein structure was generated using the DAMMIN programme (Svergun, 1999). Ten models were generated which were averaged to create the final DAMFILT model. The figures were prepared with UCSF Chimera (Pettersen *et al*., 2004). Dr. Takashi Ochi helped in processing and analysing the data.

4.4.2 Expression and Purification of Isotope-Labelled, Full-Length Orthologue of Rv3030 from *M. smegmatis*

The recombinant orthologous protein of Rv3030 from *M. smegmatis* with N-terminal, cleavable SUMO tag was isotope-labelled, and purified from the soluble fraction. Two

batches of protein were labelled, expressed and purified. In one of the batch, the protein had ^{15}N isotope labelling, and in the other batch, the protein was double labelled with ^{15}N and ^{13}C isotopes.

Culture growth and Expression: A loop of frozen BL21 cells of the expression clone was plated on LB-agar-kanamycin plate and incubated overnight at 37°C. All the cells were collected from the plate and inoculated in 250 ml LB media with kanamycin at final concentration of 30 µg/ml. The culture was grown overnight at 37°C, with constant shaking at 220 rpm. The culture was scaled up to 4 litres by adding 30 ml of the overnight culture in each of the fresh 1-litre LB media with kanamycin. Cells were grown at 37°C till OD_{600} reached 0.78, and the culture was spun (4200 x g, 4°C, 15 min). The pellet was washed by re-suspending it in 1 X M9 salts and centrifuging (5000 x g, 4°C, 10 min). This wash step was repeated four-times to remove any residual LB media, before the expression of the isotope-labelled protein.

An expression media was prepared as described: 1X M9 salts, 2 mM $MgSO_4$, 50 µM $CaCl_2$, 0.4% unlabelled glucose or 6 g/L ^{13}C glucose (Sigma), 1 g/L $^{15}NH_4Cl$ (Sigma), 30 µg/ml kanamycin, 1 X micronutrients solution and 1 X vitamins solution (see appendix-VI for composition of stock solutions of micronutrients and vitamins). The washed pellet was re-suspended in 1-litre expression media, and the cells were allowed to recover by incubating at 20°C for 1.5 hours with constant shaking. The culture was induced with 0.5 mM IPTG for protein expression, and grown overnight (14-16 hours) at 20°C with constant shaking.

Purification: The purification protocol of labelled protein was the same as that for the unlabelled protein, described in chapter-3, section 3.4.4.1.

4.4.3 NMR Spectroscopy of Apo Full-Length Orthologue of Rv3030 from *M. smegmatis*

4.4.3.1 Sample Preparation, Data Collection and Processing of Apo-Protein

The sample (600 µl) for NMR measurements contained 200 µM protein, 10% D_2O (Sigma) in 50 mM NaCl and 20 mM sodium phosphate pH 6.5, and 1 mM DTT. Samples were placed in 5 mm NMR tubes for data acquisition. All spectra were recorded at 298 K on Bruker Avance III AV600 spectrometer equipped with HCNF probe and actively-shielded *z*-gradients, by Dr. Katherine Stott. The expriments recorded are listed in **Table 4.1**. The backbone experiments were acquired using non-uniform sampling and processed according to a compressed sensing protocol obtained from Dr. Mark Bostock (personal communication, Dr. Katherine Stott).

Experiment	Nucleus	Field Strength (MHz)	Water Supprssion	Data Processing
HSQC	^{1}H, ^{15}N	600	WATERGATE	FT
HSQC	^{1}H, ^{13}C	600	Presaturation	FT
HNCA	^{1}H, ^{15}N, ^{13}C	600	WATERGATE	CS
HN(CO)CA	^{1}H, ^{15}N, ^{13}C	600	WATERGATE	CS
HNCACB	^{1}H, ^{15}N, ^{13}C	600	WATERGATE	CS
HN(CO)CACB	^{1}H, ^{15}N, ^{13}C	600	WATERGATE	CS
HNCO	^{1}H, ^{15}N, ^{13}C	600	WATERGATE	CS
^{15}N-NOESY-HSQC	^{1}H, ^{15}N	600	WATERGATE	FT
^{15}N-TOCSY-HSQC	^{1}H, ^{15}N	600	WATERGATE	FT
^{13}C-NOESY-HSQC	^{1}H, ^{13}C	600	Presaturation	FT
HCCH-TOCSY	^{1}H, ^{13}C	600	Presaturation	FT
(H)CC(CO)NNH-TOCSY	^{1}H, ^{15}N, ^{13}C	600	WATERGATE	CS
H(CCCO)NNH-TOCSY	^{1}H, ^{15}N, ^{13}C	600	WATERGATE	CS

Table 4.1 NMR experiments recorded on apo orthologue of Rv3030 from *M. smegmatis* in apo-state.

FT: Fourier transformation; CS: compressed sensing

4.4.3.2 Resonance Assignment and Restraint Generation

The resonance assignment of the apo-protein was done using CCPNMR Analysis software. Backbone resonance assignments were obtained from multidimensional HNCO, HNCA, HN(CO)CA, HNCACB and HN(CO)CACB experiments, and also from the NOESY and TOCSY spectrum. Sidechains were assigned using various 3D TOCSY experiments: ^{15}N-TOCSY-HSQC, HCCH-TOCSY, (H)CC(CO)NNH-TOCSY and H(CCCO)NNH-TOCSY. Dr Katherine Stott taught and helped in doing the assignments. The dihedral angles were calculated using DANGLE (Cheung *et al*., 2010), and ambiguous distance restraints were obtained by peak-picking the 3D NOESY spectra, by Dr. Katherine Stott.

4.5 Results

4.5.1 SAXS analysis of Full-Length Orthologue of Rv3030 from *M. smegmatis*

A SEC-SAXS study was carried out to investigate about the overall size and shape of the full-length orthologue of Rv3030 in *M. smegmatis.* The protein eluted as a single peak from the Superdex-200 column, immediately prior to X-ray irradiation. The data frames were collected for the peak and were averaged to generate the scattering intensity plot **(Figure 4.3 A)**.

The Guinier plot of ln (I(q)) against q^2 was a perfectly straight line, demonstrating that the sample was not aggregated **(Figure 4.3 B)**. The value of the radius of gyration calculated from the Guinier analysis is 20.25 ± 0.04 Å. The qRg limits used to determine the value of Rg from the plot were 0.533 to 1.18. The Kratky profile is bell-shaped **(Figure 4.3 C)** indicating that the protein is folded, as observed from the CD spectrum mentioned in chapter-3. Furthermore, the pair-wise distribution function (P(r)) is a single, bell-shaped curve, which is the characteristic of a globular protein **(Figure 4.3 D)**. This is consistent with the results of SV-AUC (in chapter-3), thereby, confirming that the protein exhibits a globular structure. The calculated D_{max} is 68.87 Å.

The *ab initio* shape of the protein was constructed from the SAXS data. A total of 10 models were generated using DAMMIN, and the model scattering curves fitted well with the

experimental scattering curve, with *Chi*^{2} values between 3.42 and 6.07. The models were reasonably globular with a bulge at one side. The average features of the 10 models are represented in the single DAMFILT model, with mean NSD value of 0.521 **(Figure 4.4)**.

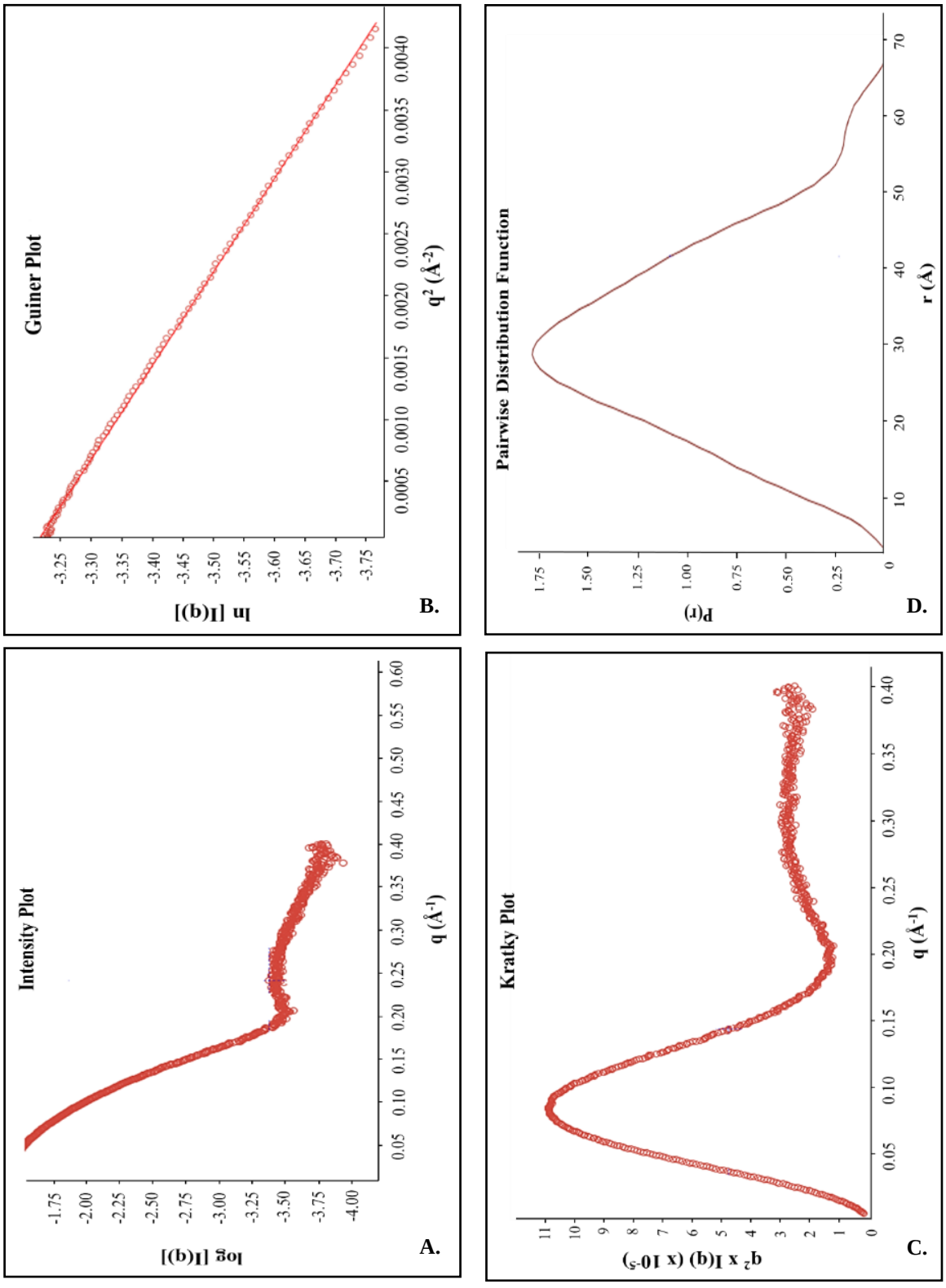


FIGURE 4.3 SAXS data profiles for full-length orthologue of Rv3030 from *M. smegmatis*.

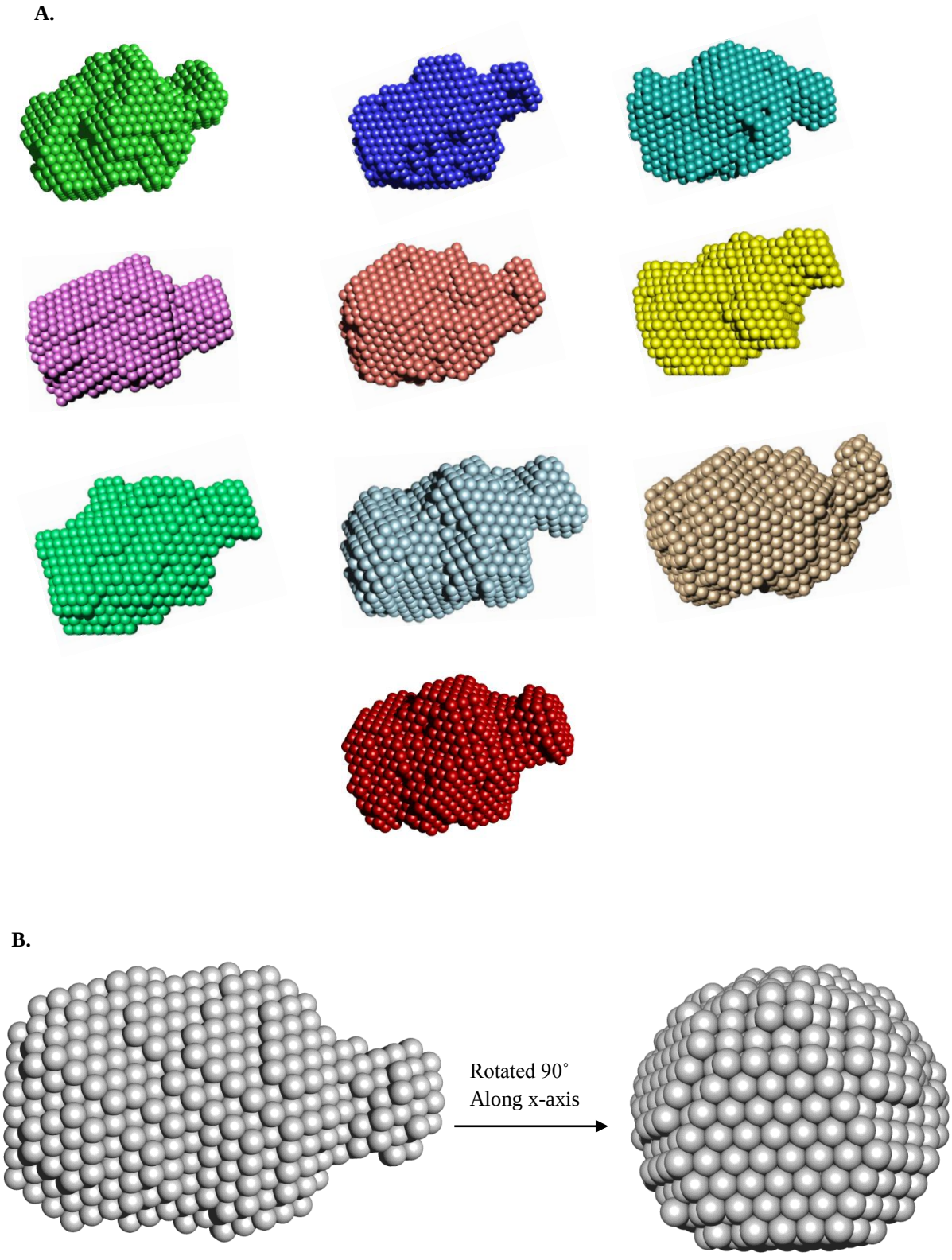


FIGURE 4.4 ***Ab initio*** **shape reconstruction of orthologue of Rv3030 from *M. smegmatis.***

A. 10 DAMMIN models constructed from experimental scattering curves.
B. Averaged DAMFILT model, showing the overall shape of the enzyme. Molecular surface representated as spheres.

4.5.2 Expression and Purification of Isotope Labelled, Full-Length Orthologue of Rv3030 from *M. smegmatis*

High-levels of the ^{15}N labelled and double-labelled (^{15}N, ^{13}C) recombinant proteins with cleavable SUMO tag at the N-terminus were purified to homogeneity by affinity chromatography and the tag was subsequently cleaved. The trace from the gel-filtration run showed that the tag-less, labelled proteins behaved as a monomer in the solution, similar to the unlabelled protein **(Figure 4.5 and Figure 4.6)**.

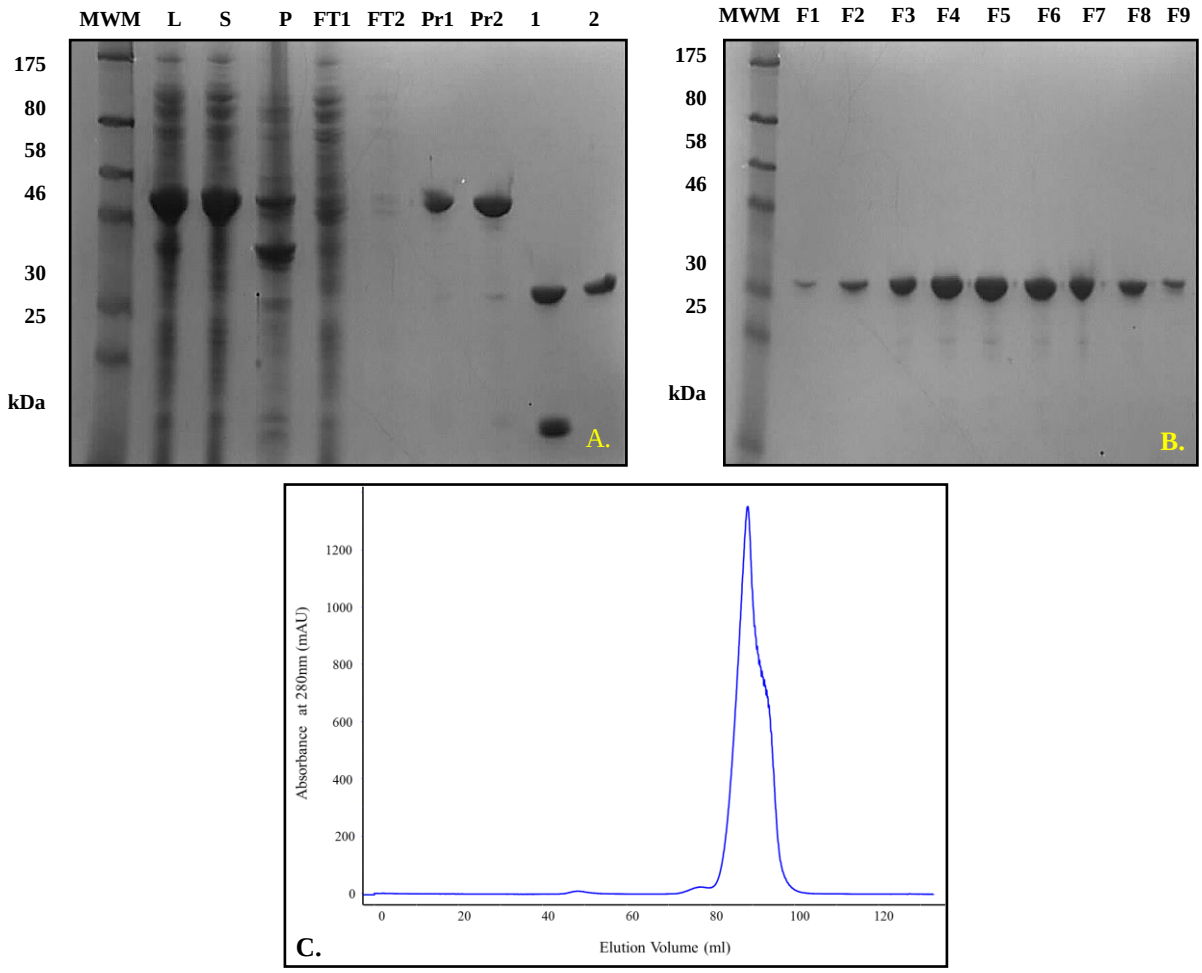


FIGURE 4.5 Expression and purification of ^{15}N labelled orthologue of Rv3030 from *M. smegmatis*.

A. Metal affinity purification of labelled, SUMO tagged protein on Ni-sepharose column, cleavage of SUMO tag by UlpI protease and reverse binding on Ni-sepharose column. (MWM: molecular weight marker, L: cellular lysate, S: supernatant, FT1 and FT2: flow through, Pr1 and Pr2: tagged protein, 1: protein after Ulp1 cleavage of SUMO tag, 2: tag less protein obtained in the flow through after reverse binding).

B. Gel filtration of tag-less labelled protein on Superdex-200 column. (MWM: molecular weight marker, F1 to F9: protein fractions).

C. The gel filtration profile.

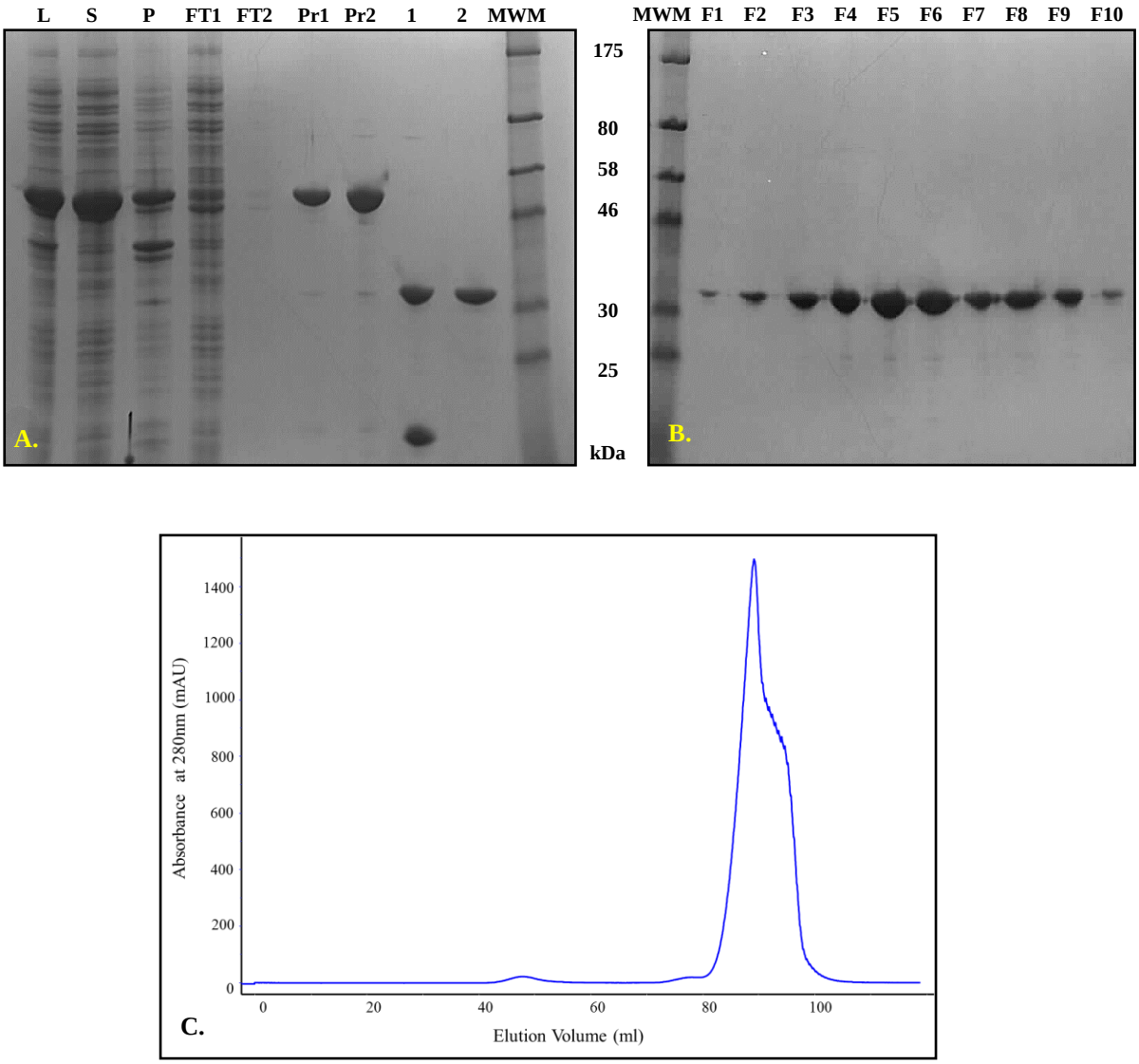


FIGURE 4.6 Expression and purification of ^{15}N and ^{13}C labelled orthologue of Rv3030 from *M. smegmatis.*

A. Metal affinity purification of labelled, SUMO tagged protein on Ni-sepharose column, cleavage of SUMO tag by UlpI protease and reverse binding on Ni-sepharose column (MWM: molecular weight marker, L: cellular lysate, S: supernatant, FT1 and FT2: flow through, Pr1 and Pr2: tagged protein, 1: protein after Ulp1 cleavage of SUMO tag, 2: tag less protein obtained in the flow through after reverse binding).

B. Gel filtration of tag-less labelled protein on Superdex-200 column.
(MWM: molecular weight marker, F1 to F10: protein fractions).

C. The gel filtration profile.

4.5.3 Structural features of Apo Full-Length Orthologue of Rv3030 from *M. smegmatis* determined by NMR Spectroscopy

4.5.3.1 ^{15}N HSQC Spectrum

The NMR measurements were made on protein at a concentration of 200 µM protein, in a buffer of 50 mM NaCl and 20 mM sodium phosphate pH 6.5 that contained 10% D_2O and 1 mM DDT. The ^{1}H-^{15}N HSQC spectrum of the protein showed that the peaks were well dispersed, in general had relatively narrow linewidths for a protein of this size (28 kDa), and were not severely overlapped **(Figure 4.7)**. This confirmed that the protein was folded, and adopted a stable monomeric conformation under these conditions.

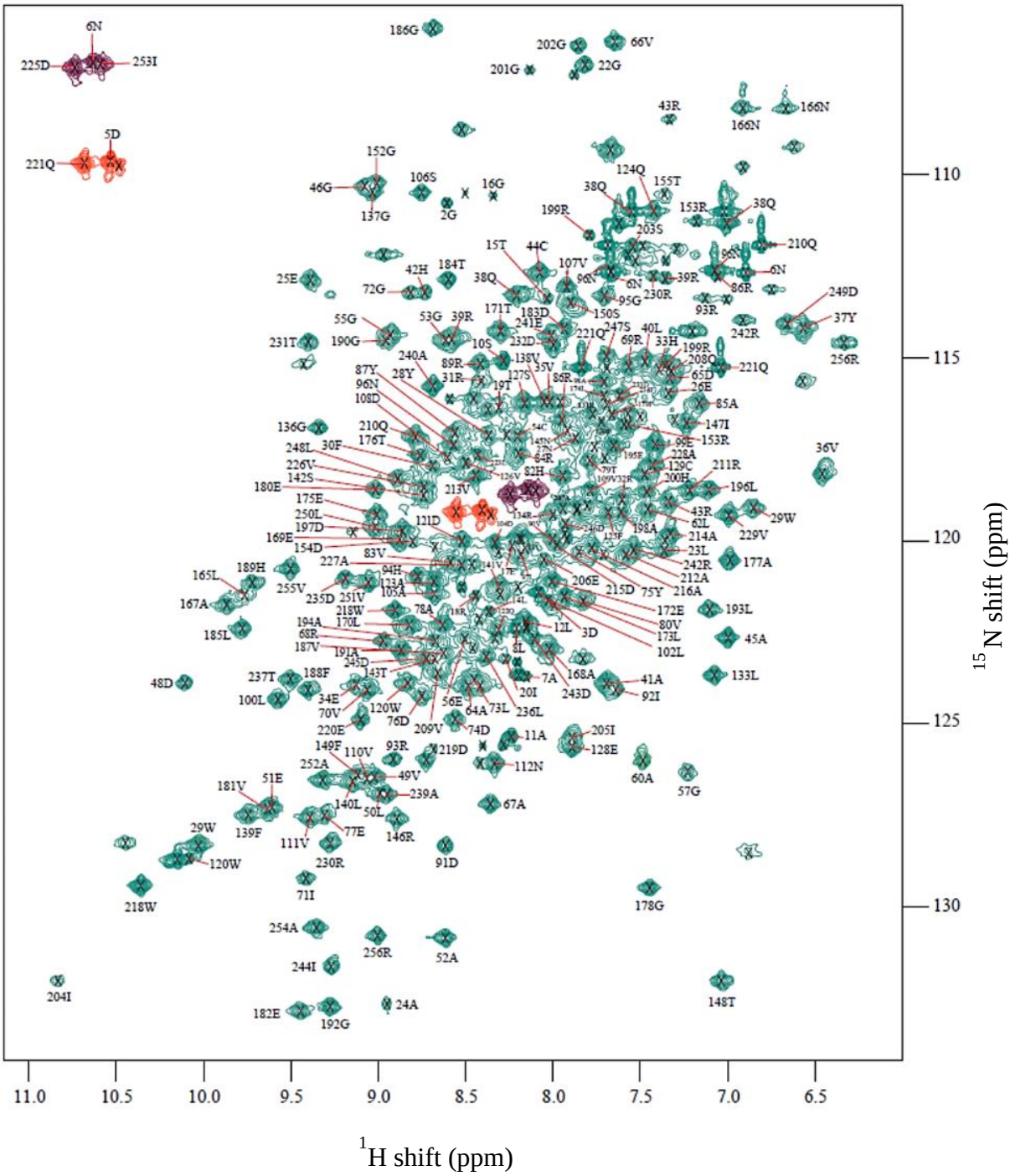


FIGURE 4.7 The ^{15}N HSQC spectrum of apo-orthologue of Rv3030 from *M. smegmatis.*

^{1}H-^{15}N HSQC spectrum of apo-orthologue of Rv3030 from *M. smegmatis* annotated with assignments of backbone and sidechain amides. Assignments of the peaks coloured purple and orange are shown on top left corner.

4.5.3.2 Resonance Assignment

The resonance assignment was completed for the all the residues except for 113-115, 117-118, and 156-164. The sidechain assignment was done using various 3D TOCSY experiments **(Figure 4.8)**. The backbone resonances were assigned using a sequential assignment strategies based on ^{1}H, ^{15}N and ^{13}C nuclei **(Figure 4.9)**. The sequential assignment process was interrupted by presence of 15 prolines residues in the protein sequence, which was bridged by Dr. Katherine Stott, using connections apparent in the NOESY spectra.

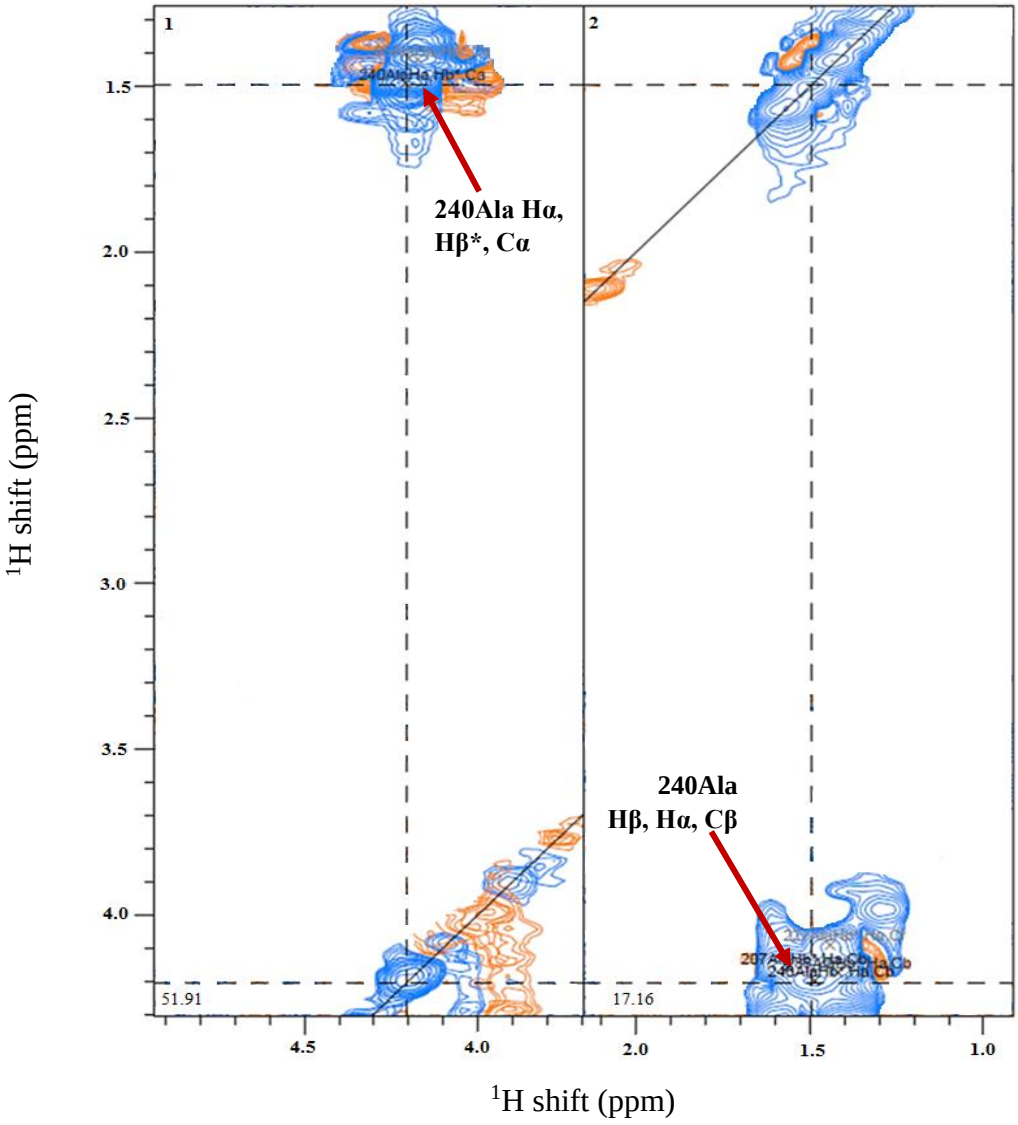


FIGURE 4.8 The HCCH-TOCSY spectrum, used for sidechain assignment.
Representative 2D 1H-1H planes for Ala240. The ^{1}H-^{13}C-^{13}C-^{1}H cross peaks of the assigned residue are marked with red arrows.

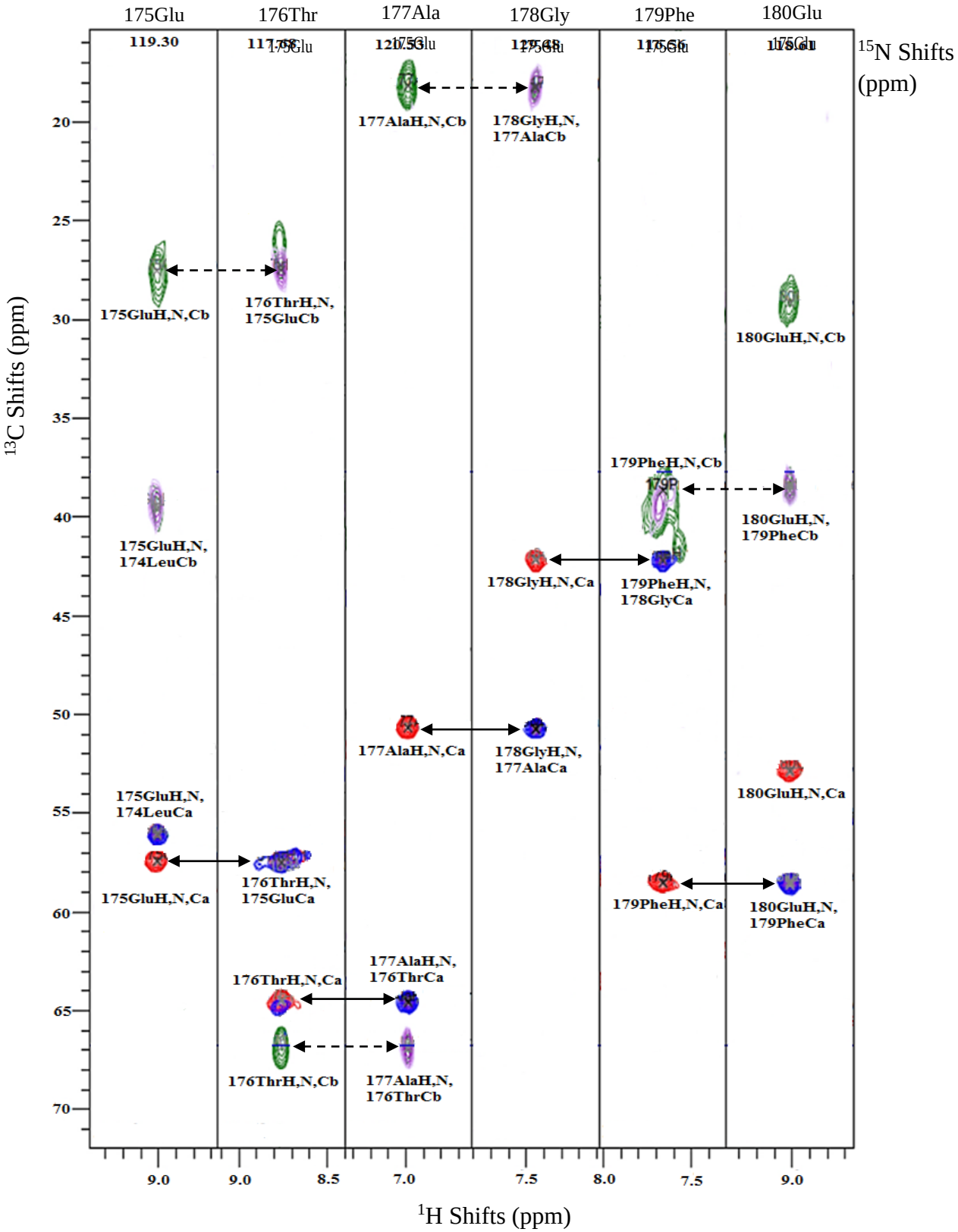


FIGURE 4.9 Sequential backbone assignments.

A stretch of six residues (175Glu-180Glu) of the orthlogue of Rv3030 from *M. smegmatis*. The sequential connections between the ^{13}Cαs of the residues are marked with black arrows between HNCA (red) and HN(CO)CA (blue) spectra. The ^{13}Cβ connections are marked with black dashed arrow between HNCACB (green) and HN(CO)CACB (purple) spectra.

4.5.3.3 Secondary Structure Characterisation and Protein Dynamics

The secondary structures of Rv3030 orthologue in *M. smegmatis* were calculated using DANGLE, which gave a measure of backbone Phi and Psi angles **(Figure 4.10)**. The measured secondary structures are mapped onto the sequence alignment between the *M. tb* and *M. smegmatis* protein **(Figure 4.11)**.

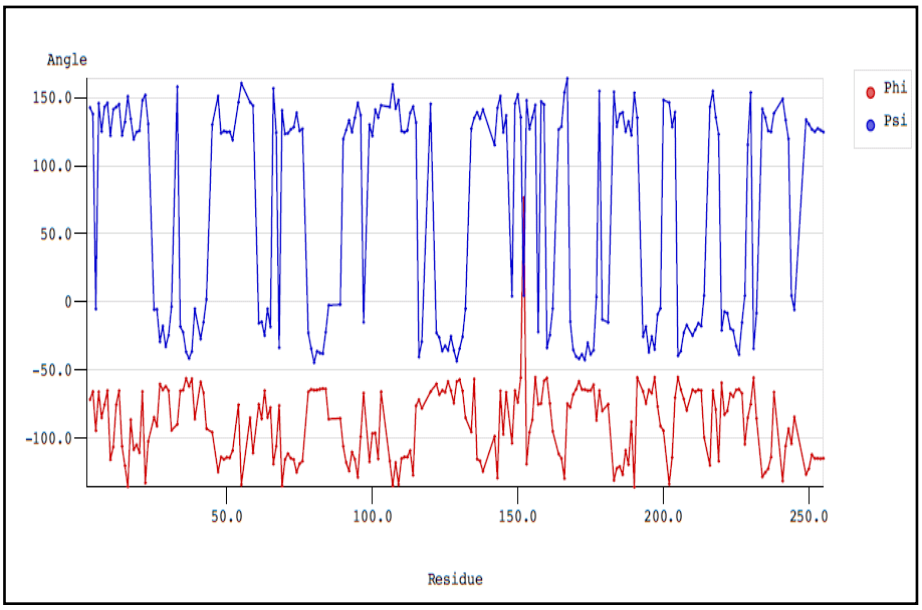


FIGURE 4.10 Measured torsion angles of the backbone of orthologue of Rv3030 from *M. smegmatis.* Angles calculated from DANGLE.

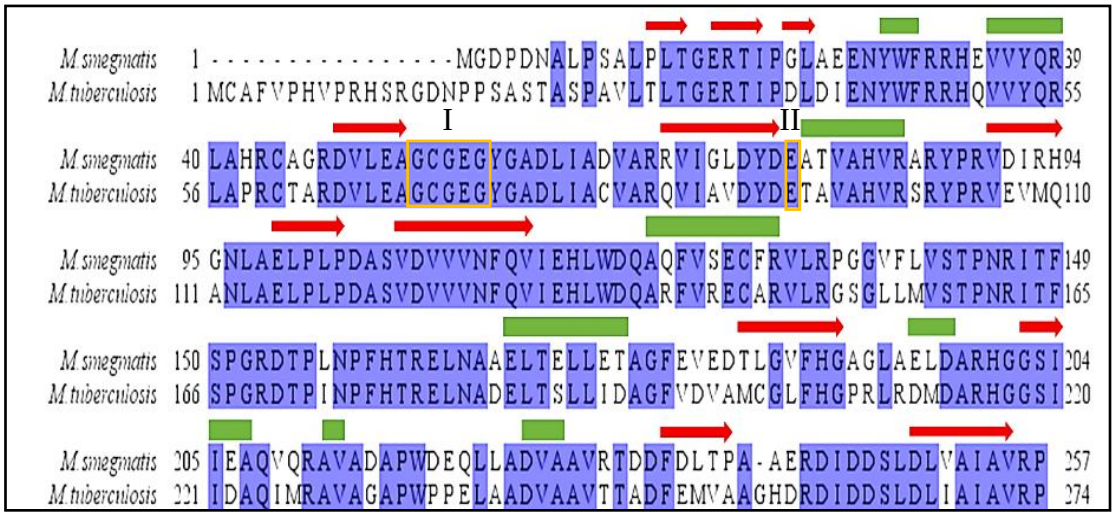


FIGURE 4.11 Sequence alignment between Rv3030 and its orthologue from *M. smegmatis* annotated with secondary structures calculated from DANGLE.

Secondary structures are annotated as follows- α-helix: green rectangles, β-strands: maroon arrows. Motif- I and II of SAM-MTs are marked with yellow box.

The dynamics of apo-protein were measured using ^{1}H-^{15}N heteronuclear NOE (HetNOE). The results showed that, using a cut-off value of 0.6, residues 24-256 were relatively rigid. However, the residues 1-23 at the N-terminal of the protein showed a substantial decrease in the HetNOE values as compared to the rest of the protein, indicating high flexibity of the N-terminus of the protein **(Figure 4.12)**.

The cross-peaks in 3D NOESY were assigned by Dr. Katherine Stott, and validated by presence of a return cross peak, where possible. Consistent stretches of long-range NOEs were observed between the β-strands. Using this information, the core β-sheet of the protein could be mapped out. There are 8 β-strands, of which two strands are anti-parallel, arranged in a topology typical of the Rossmann fold **(Figure 4.13)**.

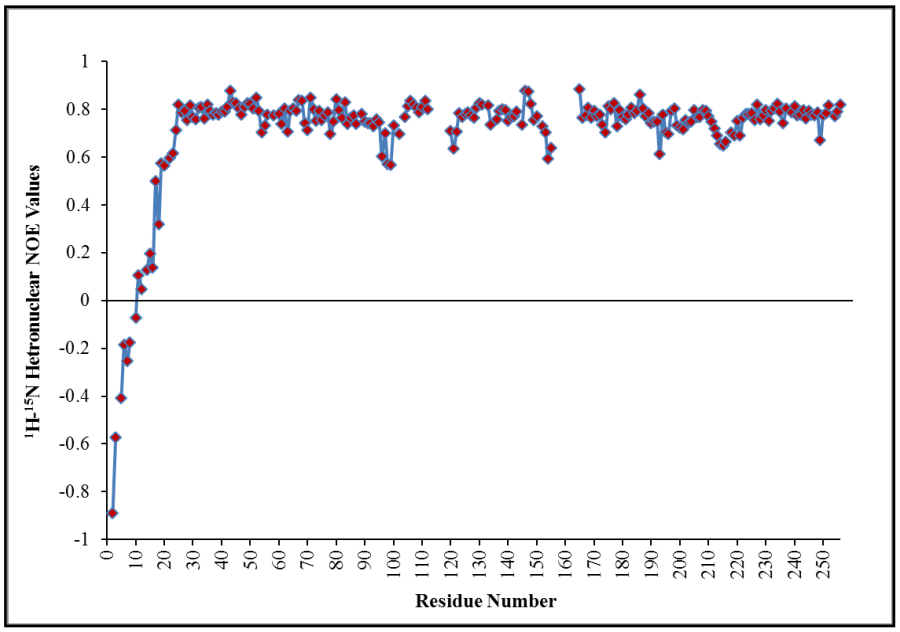


FIGURE 4.12 ^{1}H-^{15}N heteronuclear NOEs for apo-orthologue of Rv3030 from *M. smegmatis.*

The decrease in the values of the heteronuclear NOE for residues 1-23 in the protein indicates that this region is disordered.

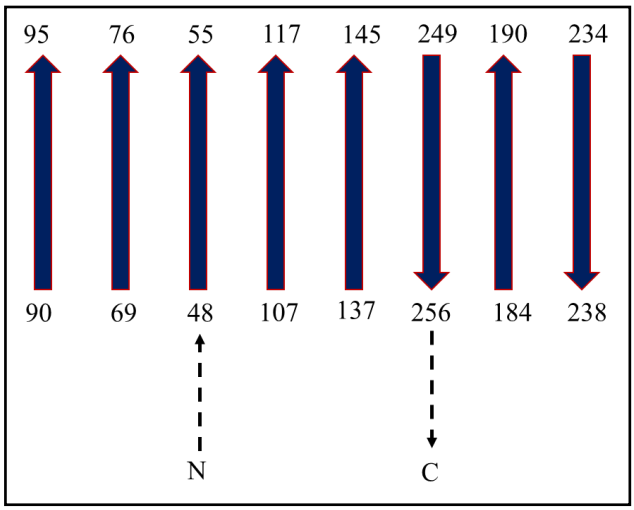


FIGURE 4.13 β-sheet topology of orthologue of Rv3030 from *M. smegmatis.*

The β-sheet topology of the enzyme was mapped using the calculated secondary structures and long range NOEs. Numbers represent the residue range. N: N-terminus, C: C-terminus.

4.5.3.4 Homology model of orthologue of Rv3030 in *M. smegmatis* based on NMR data

In the light of experimentally derived information about the structural features of the enzyme, the three-dimensional structure was re-modelled. Distant homologues of known structures were used as the structural templates to construct a model of a large part of the structure of the *M. smegmatis* orthologue of Rv3030. The details of the templates used for model building is given in **Table 4.2.**

The homology model shows that there are 7 β-strands, arranged in the typical Rossmann fold topology **(Figure 4.14 A)**. The region between the residues 190-240 was modelled with low confidence, as it was the region with lowest identity and greatest structural divergence among the templates used for model building. It might be possible that the loop in this region (marked with a dashed box in **Figure 4.14 A**) could possibly form the additional anti-parallel β-strand of the β-sheet, as calculated from the NMR data. Supportively, the overall shape of the homology model of the enzyme exhibits high similarity with the *ab initio* shape reconstituted for the enzyme from the SAXS data **(Figure 4.14 B)**.

PDB ID	ORGANISM	SEQUENCE IDENTITY %
4INE	*Caenorhabditis elegant*	21
4KRH	*Haemonchus contortus*	20
5EGP	*Aspergillus fumigatus*	18
1VL5	*Bacillus halodurans*	20
4P7C	*Pseudomonas syringae*	19
4QNU	*Escherichia coli*	18
3DLI	*Archaeoglobus fulgidus*	20
2P7I	*Pectobacterium atrosepticum*	18
4HTF	*Escherichia coli*	20

TABLE 4.2 Structural templates used for homology modelling of orthologue of Rv3030 from *M. smegmatis.*

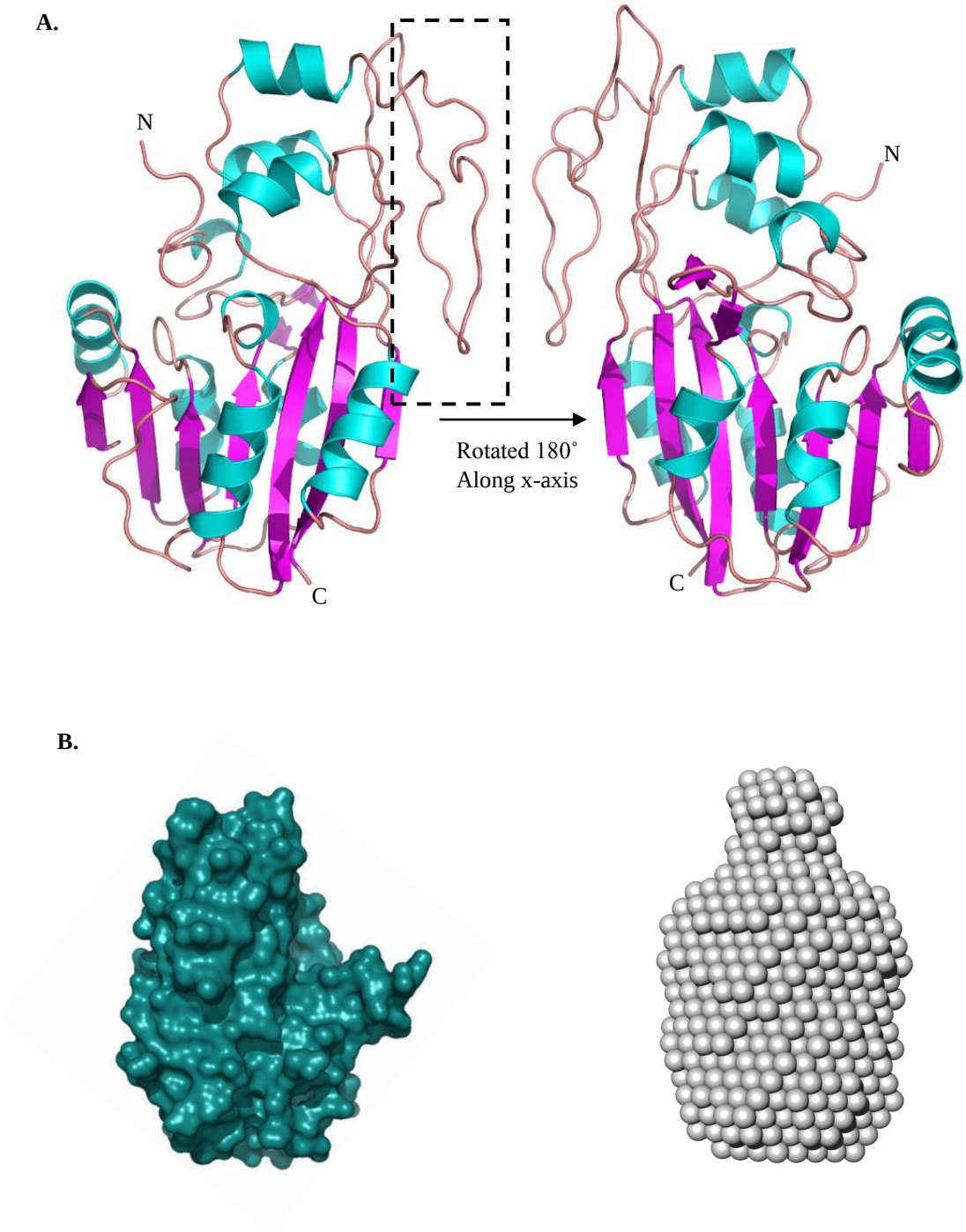


FIGURE 4.14 ***In silico* modelling of three-dimensional structure of orthologue of Rv3030 from *M. smegmatis***

A. Homology model of Rv3030 orthologue from *M. smegmatis* based on the findings on secondary structures of the enzyme determined by NMR.

B. Homology model representated as the molecular surface (left) and the *ab initio* model from the SAXS study (right), shown in the same orientation. Both the models show high similarity in the overall shape of the enzyme.

4.6 Discussion

The results in this chapter show that NMR and SAXS are viable methods to gain insights about overall structural properties of the enzyme, using orthologue of Rv3030 from *M. smegmatis* as the representative system. The analysis from SAXS demonstrated that the protein was folded and was globular in shape, with a bulge on one end. Initial investigation to determine the suitability of protein to carry out NMR-based structure determination was assessed by the ^{15}N-HSQC spectrum, which showed that the peaks were well-dispersed. Based on this, further expression and purification of protein labelled with ^{15}N and ^{13}C isotopes was done to elucidate its three-dimensional structure. Although *M. smegmatis* orthologue of Rv3030 is a large protein, the resonance assignment process was relatively straightforward due to well-resolved peaks, and generated unambiguous data set. The secondary structures and backbone torsion angles of the protein were calculated using DANGLE. Long-range NOEs were observed for the 8 β-strands in the protein, which formed the core β-sheet, with a topology characteristic of the Rossmann fold. The structure calculation process is underway to determine the complete three-dimensional view of the protein's architecture.

The dynamics of this methyltransferse were studied using heteronuclear NOE, which depicted that the residues 1-23 of the N-terminus are disordered, as they had high heteronuclear NOE values. This experimental validation was reassuring, as the rationale was applied and truncations of this region were made to obtain well-diffracting crystals (discussed in chapter-3). The information gathered about the secondary structural features of the protein assisted in modelling the three dimensional structure of the protein, which exhibited the typical Rossmann fold architecture. The molecular suface of the model was similar to the *ab initio* shape reconstituted from the SAXS data.

Overall, the on-going elucidation of the overall structure of the protein will be useful in understanding the basic molecular mechanisms, such as the mode of catalysis and substrate recognition, and will provide a structural scaffold for carrying out structure-based drug designing.

CHAPTER-5

Conclusions and Future Work

Given the central role of glycobiology to the cell wall of mycobacteria, the structure, biosynthesis and biology of saccharides have become central to the design of antimicrobials. The advantage of targets that are enzymes involved in biosynthesis has been further underlined by their relative accessibility to small molecules that might be effective in killing these important agents of infectious disease. In this thesis, I describe the fundamental biochemical and structural biology studies that provide the groundwork for assessing the tractability of two enzymes for structure-guided, fragment-based discovery of new therapeutic agents.

Mycobacterial Trehalose-6-Phosphate Synthase

Trehalose plays a pivotal role in defining the architecture of the cell wall and physiology of mycobacteria. Although *M. tb* possesses two independent routes for synthesis of trehalose, the prime OtsA-OtsB2 pathway is critical for therapeutic intervention targeting trehalose biosynthesis. This is because the inactivation of this pathway leads to pronounced growth defects and mortality of the bacilli, whereas knock-outs of the other pathway, TreyYZ, have no effects on *in vitro* growth (Murphy *et al*., 2005).

The crystal structures of *M. thermoresistibile* OtsA described in this work provide the first information on a mycobacterial OtsA, giving a structural view of the *M. tuberculosis* enzyme, as both the enzymes share a high percentage of sequence identity, and the residues essential for catalysis are completely conserved. The structures, determined in apo and ligand-bound forms, present the catalytic site in both ‘open’ and ‘closed’ conformations. The apo-structure of mycobacterial OtsA, solved at 2.18 Å resolution, was the first structure of this retaining glycosyltransferase in the apo-form, with subsequent determination of the apo-structure of OtsA from *B. xenovorans*. The protein structure belongs to the GT-B structural class of glycosyltransferases, having the characteristic two Rosmann-fold domains. As observed for other GT-B glycosyltransferases, mycobacterial OtsA also exhibits different structural states defined by ligand binding, and vital for its trehalose-6-phosphate synthase activity. *M. thermoresistibile* OtsA assembles as a tetramer in both solution and in the crystal structures obtained. However, this tetrameric organisation is not conserved among different

species and different oligomeric states have been reported. For example, *E. coli* OtsA exists both as a dimer and a tetramer in solution (Gibson *et al*., 2002a).

Mycobacterial OtsAs can process different NDP-glucose donor-substrates (Pan *et al*., 1996, 2002). The biochemical data of *M. thermoresistibile* OtsA showed that the enzyme has a predisposition for purine-derived NDP-glucoses, particularly for ADP-glucose. This is in contrast to *E. coli* OtsA, which prefers UDP-glucose as the donor substrate (Gibson *et al*., 2002b). Important insights are gained from the donor-bound structures of *M. thermoresistibile* OtsA, which exemplifies structural differences that could potentially contribute to different glucose-donor substrate preferences. The structural environment of the nucleotide-binding pocket defines the preference for ADP-glucose by *M. thermoresistibile* OtsA. The substitution of residues of the catalytic site (Ile295 and Phe399 in *E. coli* for Leu319 and Val363 in *M. thermoresistibile,* respectively) enlarges and deepens the catalytic site cleft, allowing the primary amine of the adenine ring to occupy a deep pocket, which might account for the strong preference for ADP-glucose by mycobacterial OtsAs. In support of this hypothesis, analysis of amino-acid conservation demonstrates that, although the catalytic core of the enzyme is conserved, a few residues of the donor-binding site, especially those interacting with the nitrogenous base of the donor-substrate, are not under the pressure of evolutionary conservation. These observations give clues as to how a sequence-structure relationship alters the substrate specificity.

An early-stage drug discovery exercise, using a fragment-based approach, was carried out against this target. The screening of the fragment library by a fluorescence-based thermal-shift assay led to identification of hits that bind to the enzyme, and have an inhibitory effect on the activity. However, various attempts to obtain fragment-bound crystal structures did not bear any fruit. The preliminary data on the fragments identified may serve as a starting point for potential structure-guided drug-designing campaign against this vital target.

Overall, the structure-function analyses of *M. thermoresistibile* OtsA described in this thesis provide an in-depth knowledge of the mycobacterial enzyme. This knowledge, along with the preliminary data on prospective fragment hits, opens the avenues for further investigation of this target at the molecular level, and for rational development of specific inhibitors against it.

The structural basis for donor-substrate specificity of OtsA presented in this thesis requires further experimental scrutiny. This could be achieved by creating mutations of the interacting

residues, particularly Leu319 and Val363, in the catalytic site. The effects on the structure of the enzyme can be studied by obtaining structures of the mutants bound with various glucose-donor substrates, and important insights on how the catalysis is modulated, can be gained through the kinetic studies. The structures and functional assay of the wild type mycobacterial OtsA reported in this study can serve as a reference for the comparative analyses.

Additionally, it is of great interest to understand whether formation of a tetrameric assembly by mycobacterial OtsA has any biological relevance, as it is not conserved among species. It is also of interest to investigate whether OtsA in mycobacteria forms any complex with OtsB2, the phosphatase in the pathway. The rationale for this is derived from the fact that in some organisms, such as *S. cerevisae*, OtsA and OtsB are part of a complex for trehalose biosynthesis (Bell *et al.*, 1998). This is especially important for the drug-development targeting trehalose biosynthesis, as inhibitors might then be developed that target the entire complex and more effectively inhibit trehalose biosynthesis in *M. tb*, thereby seriously damaging the cell wall formation and viability of *M. tb*.

The identification of chemical moieties that bind to and inhibit this mycobacterial OtsA has laid the foundation for initiating drug-discovery campaign against this target. This will provide a completely novel spectrum to combat the TB infection by targeting an indispensible saccharide in the context of biology of *M. tb*. Although it was not possible to obtain structural information with the current set of fragments tested, it would still be of interest to further screen these molecules by other biophysical methods, such as ITC or surface plasmon resonance to obtain their binding affinities. These fragments can then be modified by addition or substitution of different functional groups, and their potency can be measured. Alternately, virtual screening can also help in identification of analogues of current fragment hits or new chemical scaffolds can be identified. The structures of *M. thermoresistibile* OtsA can serve as structural templates for performing the *in silico* identification of hits, which can be subsequently validated and optimised experimentally.

Rv3030: SAM dependent-methyltransferase in MGLP biosynthesis

Polymethylated polysaccharides (PMPSs) are unique to mycobacteria, and of its two types, only MGLPs are found in the slow growing pathogenic mycobacteria. This makes this pathway interesting to study, especially in determining the structures of the essential enzymes involved in its biosynthesis and in designing a drug-development strategy against them. Of the many enzymes involved in this pathway, Rv3030, a SAM-dependent methyltransferase, is a critical target, as demonstrated by earlier knock-out studies of this enzyme, which showed low-levels of MGLP production and severe impaired growth of the bacteria. In the current study, this target was structurally evaluated using a combinatorial approach of various structure-determination techniques, and is a good example of how small contributions from different techniques can complement each other effectively to gain insights into an enzyme's structural features.

The initial investigation by crystallisation was not successful, as low-resolution diffraction crystals of the full-length orthologue of Rv3030 from *M. smegmatis* were observed only in one condition of the many that were trialled, and these took a long time to form and the condition was not reproducible. Data from circular dichroism indicated that the full-length protein was folded. This was reassuring in the absence of a biochemical assay to test the activity of the enzyme, due to lack of availability of its required substrate. The homology model of the structure of the protein indicated that the N-terminus of the protein was probably disordered, potentially interfering with crystallisation. These N-terminus residues were, therefore, truncated in order to make the protein ammeneable for crystallisation and several constructs were prepared with varying degrees of truncations. Although, there were no instability issues with the truncated proteins, and all the truncated proteins expressed in exceptionally high amounts with high purity, no crystals were produced.

The analyses by AUC and SAXS demonstrated that the protein was monomeric, folded and globular. The overall shape of the protein conformed to a slightly elongated globular structure with a bulge on one end. With encouraging data about the shape and folding state of the protein, I explored NMR spectroscopy as a viable route to gain insights into the full-length structure of Rv3030 orthologue in *M. smegmatis*. Expression and purification of labelled protein was straightforward, similar to that of the unlabelled protein, with ample amount of labelled protein purified to homogeneity. The ^{15}N-HSQC fingerprint of the protein revealed

that, even though the size of the protein was on the upper limit for NMR study, the spectrum was well resolved, with few severely overlapped peaks. This encouraged me to undertake advanced multi-dimensional NMR experiments, for thorough assessment of the structure of the full-length protein. The backbone and side-chain assignments proved feasible due to well-resolved peaks in the spectra, and secondary structures were characterised. The structure calculations are still underway to gain a holistic picture of the three-dimensional structure of the methyltransferase.

Nevertheless, some important information was gained from the NMR analysis through characterisation of the secondary structure and dynamic behaviour of the protein in solution. The protein has the β-sheet topology typical of the Rossmann-fold. The experimental data from ^{15}N-heteronuclear NOEs pointed out to the N-terminal region being intrinsically disordered. This was reassuring, and experimentally validated the observation from the *in silico* model, based on which truncated constructs had been prepared.

Overall, SAXS and NMR spectroscopy have been shown to be feasible approaches for gaining structural insights into this protein. As many experienced structural biologists can attest to, these methods have been vital tools in gaining knowledge about intrinsically disordered proteins. The elucidation of complete three-dimensional structure of Rv3030 will answer some key questions. Firstly, it would be possible to determine if the disordered N-terminus is making any interactions with rest of the structure, and if this region plays an important role in the overall protein stability. Secondly, it would also be possible to determine if this region defines the architecture of the catalytic pocket. This can be achieved by utilising the already prepared truncated constructs, and expressing and purifying the labelled, truncated proteins. The ^{15}N-HSQC spectra of the full-length and truncated proteins can then be compared and evaluated for any observeable changes in the spectra, which can give a measure on the effects that the truncations might have on the overall protein folding. This may also be useful for fine-tuning the domain boundaries of the enzyme that can be used for crystallisation. Also, it would also be possible to determine if the residues in this region are part of the binding pocket, or if they facilitate ligand binding. This can be achieved by performing titration experiments with the co-factor, SAM, and measuring the chemical shift pertubations. In the broader context, given the essentiality of this enzyme in the bacilli, it might be useful to identify lead chemical molecules that target this enzyme, for the development of novel anti-TB therapies.

A final word

It is essential to understand the basic molecular mechanisms of the potential drug-targets, to exploit their full potential in effective, targeted treatment of tuberculosis. A combined structural and functional approach should provide a comprehensive understanding and pave way for the development of specific and potent anti-TB drugs.

APPENDICES

Appendix-I Chemical Shift Assignments of Apo Rv3030 Orthologue from *M. smegmatis*

The chemical shift assignments of the full-length apo Rv3030 orthologue in *M. smegmatis* (**Chapter-4, Section 4.5.3.2)** are listed in this appendix. All shifts are reported in ppm.

RESIDUE NUMBER	RESIDUE	H	N	C	CA	OTHERS
1	**Met**	-	-	174.58	54.46	4.47:Ha 2.09:Hb2 2.09:Hb3 2.62:Hg2 2.62:Hg3 2.12:He* 30.51:Cb 30.62:Cg 14.91:Ce
2	**Gly**	8.61	110.78	171.69	43.12	3.90:Haa 4.04:Hab
3	**Asp**	8.09	121.43	-	50.71	4.90:Ha 2.61:Hba 2.77:Hbb 38.75:Cb
4	**Pro**	-	-	174.99	61.72	4.45:Ha 2.34:Hb2 2.34:Hb3 1.96:Hba 2.34:Hbb 2.04:Hg2 2.04:Hg3 3.76:Hda 3.87:Hdb 30.06:Cb 25.46:Cg 49.05:Cd
5	**Asp**	8.41	119.3	174.21	52.7	4.59:Ha 2.64:Hba 2.71:Hbb 39.09:Cb
6	**Asn**	8.14	118.58	172.67	51.27	112.67:Nd2 4.70:Ha 2.80:Hb2 2.80:Hb3 2.75:Hba 2.83:Hbb 6.90:Hd2a 7.67:Hd2b 37.15:Cb
7	**Ala**	8.2	123.66	175.48	50.49	4.33:Ha 1.39:Hb* 17.42:Cb
8	**Leu**	8.21	122.41	-	51.15	4.60:Ha 1.58:Hba 1.60:Hbb 1.68:Hg 0.92:Hda* 0.93:Hdb* 39.83:Cb 25.09:Cg 21.54:Cda 23.17:Cdb
9	**Pro**	-	-	174.96	61.45	4.44:Ha 1.95:Hba 2.31:Hbb 2.04:Hg2 2.04:Hg3 3.64:Hda 3.86:Hdb 30.12:Cb 25.66:Cg 48.83:Cd
10	**Ser**	8.28	115.11	172.21	56.39	4.39:Ha 3.87:Hb2 3.87:Hb3 61.93:Cb
11	**Ala**	8.24	125.38	-	50.22	4.38:Ha 1.34:Hb* 17.67:Cb
11	**Ala**	-	-	175.19	-	
12	**Leu**	8.15	122.32	-	51.16	4.60:Ha 1.59:Hb2 1.59:Hb3 1.58:Hba 1.60:Hbb 1.68:Hg 0.92:Hda* 0.93:Hdb* 39.84:Cb 25.09:Cg 21.54:Cda 23.17:Cdb
13	**Pro**	-	-	174.72	60.99	4.45:Ha 1.92:Hba 2.31:Hbb 2.03:Hg2 2.03:Hg3 3.64:Hda 3.86:Hdb 29.89:Cb 25.66:Cg 48.60:Cd
14	**Leu**	8.37	121.95	175.58	53.21	4.37:Ha 1.60:Hb2 1.60:Hb3 1.67:Hg 0.87:Hda* 0.92:Hdb* 40.18:Cb 25.08:Cg 21.58:Cda 23.19:Cdb

RESIDUE NUMBER	RESIDUE	H	N	C	CA	OTHERS
15	**Thr**	8.04	113.4	173.04	59.86	4.28:Ha 4.19:Hb 1.17:Hg2* 67.96:Cb 19.62:Cg2
16	**Gly**	8.35	110.57	172.04	43.3	3.82:Ha2 3.82:Ha3
17	**Glu**	8.24	120.17	174.43	54.64	4.30:Ha 1.93:Hba 2.07:Hbb 2.24:Hg2 2.24:Hg3 28.55:Cb 34.46:Cg
18	**Arg**	8.45	121.59	174.28	54.17	4.47:Ha 1.68:Hb2 1.68:Hb3 29.03:Cb
19	**Thr**	8.32	116.37	172.14	59.82	4.41:Ha 4.18:Hb 1.25:Hg2* 68.15:Cb 20.09:Cg2
20	**Ile**	8.27	123.24	-	56.55	4.46:Ha 1.79:Hb 1.12:Hg1a 1.42:Hg1b 0.86:Hg2* 0.80:Hd1* 37.12:Cb 24.54:Cg1 15.40:Cg2 10.99:Cd1
21	**Pro**	-	-	175.44	60.45	3.11:Ha 1.61:Hba 1.70:Hbb 1.76:Hga 1.93:Hgb 3.52:Hda 3.65:Hdb 30.11:Cb 25.25:Cg 48.97:Cd
22	**Gly**	7.82	107.02	171.67	44.06	3.60:Haa 3.85:Hab
23	**Leu**	7.36	120.27	175.87	51.87	4.58:Ha 1.35:Hba 1.60:Hbb 1.51:Hg 0.88:Hda* 0.91:Hdb* 41.20:Cb 25.24:Cg 22.26:Cda 22.72:Cdb
24	**Ala**	8.95	132.64	177.06	54.32	4.00:Ha 1.28:Hb* 16.50:Cb
25	**Glu**	9.39	112.88	176.15	57.26	3.38:Ha 1.85:Hb2 1.85:Hb3 1.65:Hga 2.03:Hgb 27.32:Cb 34.87:Cg
26	**Glu**	7.34	115.96	175.44	53.37	5.02:Ha 29.00:Cb
27	**Asn**	7.86	117.21	174.43	56.45	4.56:Ha 2.87:Hb2 2.87:Hb3 38.82:Cb
28	**Tyr**	-	-	173.1	-	
28	**Tyr**	8.36	117.13	-	61.18	3.96:Ha 2.94:Hb2 2.94:Hb3 35.67:Cb
29	**Trp**	6.86	119.1	175.63	57.18	128.32:Ne1 3.82:Ha 3.33:Hba 3.45:Hbb 7.77:Hd1 10.03:He1 7.66:He3 7.51:Hz2 7.13:Hz3 7.25:Hh2 27.81:Cb 126.84:Cd1 118.58:Ce3 112.96:Cz2 119.78:Cz3 122.85:Ch2
30	**Phe**	8.69	117.97	176.07	59.52	4.27:Ha 2.96:Hba 3.31:Hbb 6.56:Hd* 6.18:He* 6.42:Hz 38.27:Cb 128.78:Cd* 127.51:Ce* 125.57:Cz
31	**Arg**	8.45	116.18	174.92	54.88	3.90:Ha 2.64:Hb2 2.64:Hb3 27.44:Cb
32	**Arg**	7.6	118.91	174.87	56.02	28.15:Cb
33	**His**	7.4	115.39	175.71	54.18	4.59:Ha 3.34:Hba 3.48:Hbb 27.99:Cb
34	**Glu**	9.12	123.92	176.64	57.53	4.02:Ha 1.46:Hba 1.88:Hbb 2.17:Hg2 2.17:Hg3 27.92:Cb 34.34:Cg
35	**Val**	8.02	116.2	172.96	63.47	3.69:Ha 1.99:Hb 0.68:Hga* 1.07:Hgb* 29.37:Cb 17.40:Cga 22.12:Cgb

RESIDUE NUMBER	RESIDUE	H	N	C	CA	OTHERS
36	**Val**	6.46	118.15	174.99	65.02	3.19:Ha 1.21:Hb -0.12:Hga* 0.49:Hgb* 29.28:Cb 20.13:Cga 21.38:Cgb
37	**Tyr**	6.57	114.22	175.81	61.01	3.97:Ha 2.93:Hba 3.15:Hbb 7.11:Hd* 6.72:He* 37.77:Cb 131.37:Cd* 116.72:Ce*
38	**Gln**	8.21	113.3	178.3	57.4	111.32:Ne2 4.08:Ha 2.19:Hb2 2.19:Hb3 2.40:Hga 2.57:Hgb 7.01:He2a 7.62:He2b 27.12:Cb 31.86:Cg
39	**Arg**	8.58	114.51	175.85	55.12	84.82:Ne 4.39:Ha 1.83:Hba 1.97:Hbb 1.84:Hga 1.98:Hgb 3.22:Hda 3.38:Hdb 7.36:He 27.98:Cb 25.85:Cg 40.34:Cd
40	**Leu**	7.48	115.1	175.84	52.51	4.62:Ha 1.57:Hba 1.81:Hbb 1.57:Hg 0.75:Hda* 0.78:Hdb* 39.57:Cb 25.50:Cg 24.15:Cda 21.05:Cdb
41	**Ala**	7.7	123.85	177.78	54.65	3.81:Ha 1.55:Hb* 15.48:Cb
42	**His**	8.73	113.23	176.27	57.34	4.38:Ha 3.15:Hba 3.27:Hbb 26.54:Cb
43	**Arg**	7.34	118.95	174.37	53.3	80.48:Ne 4.12:Ha 1.86:Hb2 1.24:Hba 1.84:Hbb 0.95:Hga 1.25:Hgb 2.98:Hda 3.26:Hdb 7.34:He 26.66:Cb 25.47:Cg 38.82:Cd
44	**Cys**	8.08	112.71	172.5	56.72	4.06:Ha 2.77:Hba 3.20:Hbb 27.15:Cb
45	**Ala**	7	122.66	176.94	51.47	3.85:Ha 1.44:Hb* 16.16:Cb
46	**Gly**	9.08	110.32	172.18	44.63	3.86:Haa 4.07:Hab
47	**Arg**	-	-	-	-	
47	**Arg**	7.57	116.54	174.04	52.46	4.71:Ha 1.51:Hba 1.96:Hbb 1.40:Hga 1.61:Hgb 3.00:Hd2 3.00:Hd3 31.18:Cb 25.52:Cg 41.40:Cd
48	**Asp**	10.11	123.88	173.69	52.6	5.31:Ha 2.72:Hb2 2.72:Hb3 39.06:Cb
49	**Val**	9.03	126.45	171.96	58.85	4.95:Ha 2.08:Hb 0.84:Hga* 0.88:Hgb* 33.58:Cb 19.13:Cga 20.04:Cgb
50	**Leu**	8.99	126.9	173.2	50.6	5.55:Ha 1.31:Hba 2.02:Hbb 1.40:Hg 0.81:Hda* 0.93:Hdb* 44.03:Cb 25.58:Cg 22.07:Cda 25.34:Cdb
51	**Glu**	9.61	127.28	172.56	53.44	5.13:Ha 1.36:Hba 1.84:Hbb 1.67:Hga 2.02:Hgb 27.13:Cb 34.34:Cg
52	**Ala**	8.62	130.82	171.99	48.77	4.51:Ha 1.46:Hb* 16.31:Cb
53	**Gly**	8.62	114.55	172.14	44.48	4.03:Haa 4.38:Hab
54	**Cys**	8.2	117.18	173.27	56.46	4.45:Ha 2.36:Hba 3.00:Hbb 27.47:Cb
55	**Gly**	8.93	114.39	170.67	44.88	3.65:Haa 3.87:Hab
56	**Glu**	8.5	122.68	174.52	57.58	3.91:Ha 2.10:Hba 2.57:Hbb 2.37:Hga 2.56:Hgb 28.38:Cb 34.45:Cg

RESIDUE NUMBER	RESIDUE	H	N	C	CA	OTHERS
57	**Gly**	7.23	98.3	171.61	43.72	3.25:Haa 4.34:Hab
58	**Tyr**	7.86	119.17	175.56	58.92	36.10:Cb
59	**Gly**	7.67	109.33	172.42	43.03	1.79:Haa 2.63:Hab
60	**Ala**	7.49	125.96	177.42	52.96	3.79:Ha 1.35:Hb* 14.65:Cb
61	**Asp**	7.57	116.84	175.42	55.57	4.01:Ha 1.74:Hba 1.97:Hbb 41.50:Cb
62	**Leu**	7.46	119.11	179.89	56.26	4.00:Ha 1.58:Hba 1.83:Hbb 1.75:Hg 0.84:Hda* 0.92:Hdb* 39.73:Cb 24.71:Cg 22.29:Cda 24.20:Cdb
63	**Ile**	8.16	120.21	176.3	62.65	3.57:Ha 2.09:Hb 1.22:Hg1a 1.94:Hg1b 0.89:Hg2* 0.93:Hd1* 35.11:Cb 27.43:Cg1 17.25:Cg2 10.97:Cd1
64	**Ala**	8.49	123.95	175.78	52.43	3.92:Ha 1.38:Hb* 14.87:Cb
65	**Asp**	7.34	115.55	175.54	55.46	4.51:Ha 2.70:Hba 2.99:Hbb 39.53:Cb
66	**Val**	7.65	106.38	174.91	58.58	4.77:Ha 2.17:Hb 0.87:Hga* 0.94:Hgb* 32.31:Cb 16.07:Cga 19.85:Cgb
67	**Ala**	8.36	127.2	175.09	50.49	4.83:Ha 1.39:Hb* 19.06:Cb
68	**Arg**	8.97	122.74	175.25	57.3	4.19:Ha 1.43:Hba 1.67:Hbb 1.51:Hga 1.67:Hgb 3.17:Hda 3.31:Hdb 28.52:Cb 25.31:Cg 42.05:Cd
69	**Arg**	7.57	115.26	171.74	53.88	4.52:Ha 1.91:Hb2 1.90:Hb3 1.65:Hg2 1.65:Hg3 3.20:Hd2 3.20:Hd3 31.89:Cb 25.01:Cg 42.03:Cd
70	**Val**	9.06	124.08	172.07	59.32	4.90:Ha 1.88:Hb 0.81:Hga* 0.81:Hgb* 32.28:Cb 18.54:Cga 20.39:Cgb
71	**Ile**	9.42	129.22	173.44	57.64	4.84:Ha 1.65:Hb 1.06:Hg1a 1.39:Hg1b 0.80:Hg2* 0.79:Hd1* 38.44:Cb 25.88:Cg1 15.21:Cg2 12.45:Cd1
72	**Gly**	8.82	113.22	169.04	41.69	3.14:Haa 5.10:Hab
73	**Leu**	8.45	123.83	173.52	50.78	5.55:Ha 1.29:Hba 1.60:Hbb 1.38:Hg 0.72:Hda* 0.77:Hdb* 45.38:Cb 25.58:Cg 23.45:Cda 24.97:Cdb
74	**Asp**	8.56	124.88	172.04	52.53	4.81:Ha 2.49:Hba 2.57:Hbb 45.57:Cb
75	**Tyr**	7.85	120.3	174.06	55.16	4.96:Ha 3.09:Hba 3.37:Hbb 7.10:Hd* 6.79:He* 37.30:Cb 131.32:Cd* 116.31:Ce*
76	**Asp**	8.75	124.22	174.29	51.42	5.02:Ha 2.48:Hba 2.98:Hbb 40.99:Cb
77	**Glu**	9.3	127.49	176.48	58.42	3.54:Ha 1.88:Hb2 1.88:Hb3 1.88:Hga 1.99:Hgb 27.67:Cb 34.50:Cg
78	**Ala**	8.63	122.33	178.88	53.02	4.20:Ha 1.48:Hb* 15.96:Cb
79	**Thr**	7.79	117.77	174.1	64.41	3.94:Ha 4.42:Hb 1.08:Hg2* 66.41:Cb 20.91:Cg2

RESIDUE NUMBER	RESIDUE	H	N	C	CA	OTHERS
80	**Val**	7.83	121.64	174.94	64.86	3.45:Ha 2.15:Hb 0.88:Hga* 0.96:Hgb* 29.36:Cb 18.72:Cga 21.08:Cgb
81	**Ala**	7.8	119.05	178.62	53.08	4.03:Ha 1.49:Hb* 16.06:Cb
82	**His**	7.94	118.21	174.86	57.31	4.41:Ha 3.32:Hba 3.56:Hbb 27.47:Cb
83	**Val**	8.59	120.58	175.77	65.17	3.60:Ha 2.15:Hb 1.11:Hga* 1.22:Hgb* 30.14:Cb 20.09:Cga 21.74:Cgb
84	**Arg**	8.2	117.63	175.91	56.43	4.27:Ha 1.83:Hb2 1.83:Hb3 1.64:Hga 1.82:Hgb 3.20:Hda 3.27:Hdb 28.45:Cb 25.84:Cg 41.96:Cd
85	**Ala**	7.16	116.25	176.94	51.61	4.17:Ha 1.42:Hb* 17.28:Cb
86	**Arg**	7.95	116.7	174.18	55.05	84.79:Ne 3.96:Ha 0.82:Hba 1.10:Hbb 0.56:Hga 0.81:Hgb 2.92:Hd2 2.92:Hd3 7.06:He 29.40:Cb 24.63:Cg 40.73:Cd
87	**Tyr**	8.56	117.08	-	52.2	5.32:Ha 2.95:Hb2 2.95:Hb3 7.02:Hd* 7.08:He* 36.29:Cb 131.47:Cd* 115.78:Ce*
88	**Pro**	-	-	175.86	62.4	4.61:Ha 1.94:Hba 2.44:Hbb 2.00:Hg2 2.00:Hg3 3.30:Hd2 3.30:Hd3 30.18:Cb 25.11:Cg 48.52:Cd
89	**Arg**	8.41	115.17	174.86	55.85	4.29:Ha 1.86:Hb2 1.86:Hb3 28.91:Cb
90	**Val**	8	119.94	172.91	60.72	3.74:Ha 2.19:Hb 0.67:Hg1* 0.67:Hg2* 0.66:Hga* 0.69:Hgb* 30.48:Cb 18.84:Cga 19.07:Cgb
91	**Asp**	8.62	128.29	172.96	50.88	4.83:Ha 2.70:Hba 3.18:Hbb 37.67:Cb
92	**Ile**	7.64	124.1	171.75	55.16	4.97:Ha 1.65:Hb 1.31:Hg1a 1.52:Hg1b 1.08:Hg2* 0.87:Hd1* 38.70:Cb 28.30:Cg1 16.34:Cg2 12.58:Cd1
93	**Arg**	8.91	125.98	171.87	51.96	85.39:Ne 4.75:Ha 1.61:Hba 1.92:Hbb 1.59:Hga 1.72:Hgb 2.93:Hda 3.15:Hdb 7.13:He 30.98:Cb 24.83:Cg 41.77:Cd
94	**His**	8.77	120.97	-	52.59	30.81:Cb
94	**His**	-	-	173.8	-	5.38:Ha 3.12:Hb2 3.12:Hb3 7.02:Hd2 115.74:Cd2
95	**Gly**	7.71	113.32	167.74	43.74	3.65:Haa 3.92:Hab
96	**Asn**	8.57	117.35	173.48	50.04	112.64:Nd2 4.65:Ha 2.82:Hb2 2.82:Hb3 7.07:Hd2a 7.67:Hd2b 38.17:Cb
97	**Leu**	8.18	120.38	173.13	54.98	3.99:Ha 1.58:Hb2 1.58:Hb3 1.58:Hg 0.60:Hda* 0.72:Hdb* 38.83:Cb 26.88:Cg 21.88:Cda 23.49:Cdb
98	**Ala**	7.7	115.66	175.05	51.17	4.32:Ha 1.45:Hb* 17.59:Cb

RESIDUE NUMBER	RESIDUE	H	N	C	CA	OTHERS
99	**Glu**	7.42	117.39	173.27	52.49	4.40:Ha 1.77:Hba 1.89:Hbb 2.06:Hg2 2.06:Hg3 29.30:Cb 34.24:Cg
100	**Leu**	9.57	124.33	-	50.66	4.52:Ha 1.81:Hba 1.85:Hbb 1.52:Hg 0.72:Hda* 0.92:Hdb* 39.63:Cb 25.27:Cg 24.64:Cda 22.60:Cdb
101	**Pro**	-	-	172.99	61.03	4.80:Ha 1.70:Hba 2.18:Hbb 1.98:Hga 2.45:Hgb 3.30:Hda 3.76:Hdb 26.43:Cb 30.17:Cg 48.51:Cd
102	**Leu**	8.03	121.61	-	50.55	4.90:Ha 1.16:Hba 1.78:Hbb 1.52:Hg 0.82:Hda* 0.91:Hdb* 42.90:Cb 26.91:Cg 24.49:Cda 22.73:Cdb
103	**Pro**	-	-	173.38	59.43	4.59:Ha 1.93:Hba 2.42:Hbb 1.93:Hga 1.99:Hgb 3.53:Hda 3.82:Hdb 30.76:Cb 25.26:Cg 48.79:Cd
104	**Asp**	8.33	120.06	173.76	52.74	4.02:Ha 2.48:Hba 2.66:Hbb 37.56:Cb
105	**Ala**	8.68	121.44	173.47	50.68	3.43:Ha 1.30:Hb* 15.31:Cb
106	**Ser**	8.75	110.5	172.35	56.87	4.57:Ha 3.41:Hba 3.87:Hbb 64.09:Cb
107	**Val**	7.92	113.08	171.72	58.05	4.71:Ha 2.21:Hb 0.63:Hga* 0.78:Hgb* 33.98:Cb 21.02:Cg2 16.87:Cga 20.93:Cgb
108	**Asp**	8.61	117.75	174	54.41	4.90:Ha 2.94:Hba 3.05:Hbb 42.68:Cb
109	**Val**	7.79	118.62	173.34	57.83	5.67:Ha 1.89:Hb 0.93:Hg1* 0.93:Hg2* 35.90:Cb 19.67:Cg1 19.67:Cg2
110	**Val**	9.06	126.5	171.93	58.84	5.35:Ha 1.87:Hb 0.84:Hga* 0.95:Hgb* 32.66:Cb 20.68:Cga 20.12:Cgb
111	**Val**	9.39	127.55	172.59	59	4.28:Ha 1.35:Hb -0.25:Hga* 0.58:Hgb* 32.36:Cb 18.26:Cga 19.69:Cgb
112	**Asn**	8.34	126.08	-	51.87	5.12:Ha 2.75:Hba 3.06:Hbb 41.14:Cb
113	**Phe**	-	-	-	-	
114	**Gln**	-	-	-	-	
115	**Val**	-	-	-	-	
116	**Ile**	-	-	-	63.13	3.61:Ha 2.07:Hb 1.17:Hg1a 1.54:Hg1b 0.64:Hg2* 0.91:Hd1* 36.02:Cb 27.87:Cg1 14.97:Cg2 12.82:Cd1
117	**Glu**	-	-	-	-	
118	**His**	-	-	-	-	
119	**Leu**	-	-	176.59	52.11	4.61:Ha 1.34:Hba 1.77:Hbb 1.78:Hg 0.24:Hda* 0.64:Hdb* 40.72:Cb 24.00:Cg 24.00:Cda 19.93:Cdb

RESIDUE NUMBER	RESIDUE	H	N	C	CA	OTHERS
120	**Trp**	8.83	123.88	174.81	57.34	128.70:Ne1 4.65:Ha 3.19:Hba 3.46:Hbb 7.36:Hd1 10.15:He1 7.77:He3 7.58:Hz2 7.18:Hz3 7.32:Hh2 27.86:Cb 125.30:Cd1 118.71:Ce3 113.04:Cz2 120.09:Cz3 123.09:Ch2
121	**Asp**	8.52	120.01	175.04	50.67	4.99:Ha 2.53:Hba 3.08:Hbb 37.76:Cb
122	**Gln**	8.31	122.42	175.79	57.77	3.86:Ha 2.22:Hb2 2.22:Hb3 2.37:Hga 2.48:Hgb 26.65:Cb 32.95:Cg
123	**Ala**	8.68	121.11	178.48	53.41	4.15:Ha 1.50:Hb* 16.03:Cb
124	**Gln**	8.19	119.94	176.87	56.7	111.02:Ne2 4.09:Ha 2.01:Hba 2.19:Hbb 2.41:Hga 2.56:Hgb 7.02:He2a 7.55:He2b 26.18:Cb 31.60:Cg
125	**Phe**	7.68	119.23	175.66	59.3	4.36:Ha 3.04:Hba 3.38:Hbb 7.25:Hd* 7.09:He* 6.91:Hz 37.25:Cb 130.10:Cd* 128.52:Ce* 128.31:Cz
126	**Val**	8.5	117.84	176.25	65.61	3.30:Ha 2.23:Hb 1.19:Hga* 1.23:Hgb* 29.35:Cb 21.73:Cga 21.31:Cgb
127	**Ser**	8.16	116.25	175.35	60.43	4.30:Ha 4.05:Hb2 4.05:Hb3 60.87:Cb
128	**Glu**	7.89	125.6	175.88	56.56	4.32:Ha 1.90:Hba 2.52:Hbb 2.26:Hga 2.44:Hgb 29.09:Cb 36.23:Cg
129	**Cys**	7.47	118.18	172.86	61.82	3.55:Ha 1.12:Hba 2.40:Hbb 23.80:Cb
130	**Phe**	7.72	116.71	174.15	59.81	3.43:Ha 2.47:Hba 3.18:Hbb 6.86:Hd* 5.87:He* 7.10:Hz 36.83:Cb 128.38:Cd* 129.67:Ce* 131.31:Cz
131	**Arg**	7.77	116.55	177.86	57.51	3.97:Ha 1.88:Hb2 1.88:Hb3 28.49:Cb
132	**Val**	7.67	-	-	61.43	
132	**Val**	-	109.34	173.62	-	4.25:Ha 2.25:Hb 0.77:Hga* 0.85:Hgb* 29.80:Cb 17.16:Cga 20.42:Cgb
133	**Leu**	7.08	123.67	174.88	52.93	4.32:Ha 1.65:Hb2 1.65:Hb3 0.71:Hda* 0.79:Hdb* 41.33:Cb 24.64:Cda 22.74:Cdb
134	**Arg**	8	119.31	-	52.89	4.01:Ha 2.49:Hba 2.70:Hbb 27.72:Cb
135	**Pro**	-	-	176.05	62.78	4.01:Ha 1.83:Hba 2.49:Hbb 2.14:Hg2 2.14:Hg3 3.53:Hda 3.63:Hdb 29.38:Cb 26.29:Cg 48.48:Cd
136	**Gly**	9.34	116.92	172.65	43.62	3.70:Haa 4.20:Hab
137	**Gly**	9.03	110.5	169.03	43.72	3.81:Haa 4.47:Hab
138	**Val**	8.05	116.24	170.46	58.41	5.25:Ha 1.96:Hb 0.94:Hga* 1.05:Hgb* 34.45:Cb 20.12:Cga 19.95:Cgb

RESIDUE NUMBER	RESIDUE	H	N	C	CA	OTHERS
139	**Phe**	9.74	127.51	171	49.84	5.64:Ha 2.52:Hb2 2.52:Hb3 6.41:Hd* 6.87:He* 6.98:Hz 39.93:Cb 127.96:Cd* 128.42:Ce* 126.70:Cz
140	**Leu**	9.15	126.57	170.98	50.59	4.95:Ha 1.12:Hba 1.89:Hbb 1.61:Hg 0.66:Hda* 0.81:Hdb* 41.09:Cb 24.30:Cg 23.70:Cda 24.28:Cdb
141	**Val**	8.31	121.27	171.53	55.48	5.42:Ha 0.87:Hb 0.51:Hg1* 0.51:Hg2* 32.63:Cb 18.72:Cg1 18.72:Cg2
142	**Ser**	8.74	118.8	172.24	53.48	5.87:Ha 3.21:Hba 3.61:Hbb 65.75:Cb
143	**Thr**	8.67	123.24	-	57.5	68.14:Cb
144	**Pro**	-	-	174.09	61.53	
145	**Asn**	7.92	117.07	175.24	49.76	4.80:Ha 2.37:Hba 2.93:Hbb 35.83:Cb
146	**Arg**	8.89	127.59	174.57	57.3	4.18:Ha 1.43:Hba 1.67:Hbb 1.52:Hga 1.66:Hgb 3.17:Hda 3.31:Hdb 28.50:Cb 25.31:Cg 42.05:Cd
147	**Ile**	7.23	116.78	175.22	62.14	3.82:Ha 2.01:Hb 1.44:Hg1a 1.71:Hg1b 1.08:Hg2* 1.08:Hd1* 35.89:Cb 27.26:Cg1 15.45:Cg2 11.02:Cd1
148	**Thr**	7.04	104.03	174.31	59.37	4.55:Ha 4.24:Hb 1.32:Hg2* 68.35:Cb 20.12:Cg2
149	**Phe**	9.11	126.4	174.81	59.89	4.56:Ha 2.70:Hba 3.54:Hbb 40.08:Cb
150	**Ser**	7.9	113.52	-	52.92	4.87:Ha 3.95:Hb2 3.95:Hb3 62.08:Cb
151	**Pro**	-	-	177.3	61.63	4.44:Ha 1.96:Hba 2.31:Hbb 2.04:Hg2 2.04:Hg3 3.76:Hda 3.86:Hdb 29.67:Cb 26.02:Cg 48.68:Cd
152	**Gly**	9.01	110.26	172.03	44.21	3.84:Haa 4.03:Hab
153	**Arg**	7.59	116.82	174.29	52.99	83.30:Ne 4.73:Ha 1.75:Hba 2.19:Hbb 1.57:Hga 1.75:Hgb 3.05:Hda 3.33:Hdb 7.18:He 31.41:Cb 24.85:Cg 41.78:Cd
154	**Asp**	8.8	120.03	173.3	52.59	4.76:Ha 2.70:Hba 2.85:Hbb 39.67:Cb
155	**Thr**	7.42	111.01	-	56.64	4.94:Ha 4.30:Hb 1.26:Hg2* 68.22:Cb 19.85:Cg2
156	**Pro**	-	-	-	-	
157	**Leu**	-	-	-	-	
158	**Asn**	-	-	-	-	
159	**Pro**	-	-	-	-	
160	**Phe**	-	-	-	-	
161	**His**	-	-	-	-	
162	**Thr**	-	-	-	-	
163	**Arg**	-	-	-	-	
164	**Glu**	-	-	173.44	-	

RESIDUE NUMBER	RESIDUE	H	N	C	CA	OTHERS
165	**Leu**	9.76	121.5	174.58	52.49	5.06:Ha 0.73:Hd1* 0.73:Hd2* 43.61:Cb
166	**Asn**	8.37	116.42	173.71	48.58	108.17:Nd2 5.07:Ha 2.53:Hba 3.83:Hbb 6.67:Hd2a 6.91:Hd2b 37.80:Cb
167	**Ala**	9.87	121.74	178.16	54.54	3.77:Ha 1.43:Hb* 16.70:Cb
168	**Ala**	8.02	122.95	179.17	53.42	4.09:Ha 1.49:Hb* 15.92:Cb
169	**Glu**	8.88	119.94	178.38	56.42	4.08:Ha 1.93:Hba 2.32:Hbb 2.31:Hga 2.61:Hgb 28.60:Cb 34.38:Cg
170	**Leu**	8.82	122.31	175.5	56.18	4.08:Ha 1.24:Hba 2.02:Hbb 1.34:Hg 0.61:Hda* 0.69:Hdb* 39.74:Cb 24.83:Cg 24.48:Cda 19.74:Cdb
171	**Thr**	8.3	114.24	173.63	66.32	4.33:Ha 3.65:Hb 1.19:Hg2* 66.21:Cb 19.41:Cg2
172	**Glu**	8.01	121.13	177.16	57.94	4.22:Ha 2.12:Hb2 2.12:Hb3 2.21:Hga 2.39:Hgb 27.69:Cb 34.37:Cg
173	**Leu**	7.94	121.52	177.26	56.27	4.05:Ha 1.89:Hb2 1.89:Hb3 1.67:Hg 0.83:Hda* 0.87:Hdb* 40.96:Cb 25.04:Cg 24.04:Cda 22.55:Cdb
174	**Leu**	7.71	116.06	177.21	56.1	4.17:Ha 1.12:Hba 1.98:Hbb 1.99:Hg 0.55:Hda* 0.59:Hdb* 39.44:Cb 24.53:Cg 23.70:Cda 21.58:Cdb
175	**Glu**	9.01	119.3	180.28	57.46	4.60:Ha 2.02:Hba 2.19:Hbb 2.37:Hga 2.70:Hgb 27.57:Cb 35.79:Cg
176	**Thr**	8.76	117.67	173.38	64.55	3.95:Ha 4.38:Hb 1.26:Hg2* 66.78:Cb 19.73:Cg2
177	**Ala**	6.99	120.55	174.15	50.69	4.08:Ha 1.48:Hb* 18.15:Cb
178	**Gly**	7.45	129.48	173.56	42.21	3.46:Haa 3.88:Hab
179	**Phe**	7.66	116.53	173.95	58.52	4.68:Ha 2.80:Hba 2.99:Hbb 38.64:Cb
180	**Glu**	9.02	118.6	174.85	52.87	4.80:Ha 2.05:Hba 2.34:Hbb 2.24:Hg2 2.24:Hg3 29.22:Cb 34.47:Cg
181	**Val**	9.64	127.32	173.67	26.16	3.99:Ha 2.19:Hb 0.81:Hga* 0.99:Hgb* 29.26:Cb 19.76:Cga 21.05:Cgb
182	**Glu**	9.45	132.85	174.48	56.44	4.23:Ha 1.85:Hba 2.11:Hbb 1.97:Hga 2.32:Hgb 29.41:Cb 34.85:Cg
183	**Asp**	7.94	114.23	172.49	51.51	4.99:Ha 2.34:Hba 2.67:Hbb 42.79:Cb
184	**Thr**	8.6	112.86	170.39	58.72	5.03:Ha 3.87:Hb 1.07:Hg2* 68.94:Cb 19.81:Cg2
185	**Leu**	9.78	122.41	171.62	51.72	5.04:Ha 1.28:Hb2 1.28:Hb3 1.56:Hg 0.73:Hd1* 0.73:Hd2* 44.52:Cb 25.67:Cg 24.60:Cd1 24.60:Cd2
186	**Gly**	8.69	105.98	170.78	41.88	3.32:Haa 5.19:Hab

RESIDUE NUMBER	RESIDUE	H	N	C	CA	OTHERS
187	**Val**	8.63	123.09	173.36	60.09	5.04:Ha 1.82:Hb 0.81:Hga* 0.99:Hgb* 30.99:Cb 19.65:Cga 19.54:Cgb
188	**Phe**	9.4	124.08	174.63	54.45	4.96:Ha 2.62:Hba 3.26:Hbb 7.26:Hd* 42.58:Cb
189	**His**	9.72	121.14	174.26	54.61	4.62:Ha 2.98:Hba 3.35:Hbb 6.75:Hd2 7.69:He1 27.08:Cb 124.31:Cd2 134.70:Ce1
190	**Gly**	8.96	114.54	171.98	41.41	3.90:Haa 4.32:Hab
191	**Ala**	8.87	122.97	179.22	53.84	4.09:Ha 1.47:Hb* 16.39:Cb
192	**Gly**	9.28	104.71	174.99	44.79	3.83:Haa 4.06:Hab
193	**Leu**	7.1	121.9	176.2	55.46	4.01:Ha 0.79:Hba 1.44:Hbb 1.09:Hg -0.29:Hda* 0.53:Hdb* 40.66:Cb 28.62:Cg 23.13:Cda 23.13:Cdb
194	**Ala**	8.67	122.72	179.47	53.28	4.57:Ha 1.56:Hb* 15.80:Cb
195	**Glu**	7.65	117.41	176.59	57.09	4.12:Ha 2.03:Hba 2.10:Hbb 2.24:Hga 2.38:Hgb 27.64:Cb 34.34:Cg
196	**Leu**	7.11	118.6	177.67	55.66	4.07:Ha 1.38:Hba 2.15:Hbb 1.92:Hg 0.77:Hda* 0.85:Hdb* 39.61:Cb 25.01:Cg 20.74:Cda 23.35:Cdb
197	**Asp**	8.86	119.76	177.04	56.54	4.26:Ha 2.86:Hb2 2.86:Hb3 2.62:Hba 2.98:Hbb 38.86:Cb
198	**Ala**	7.61	119.22	177.57	52.85	4.01:Ha 1.49:Hb* 16.08:Cb
199	**Arg**	7.36	115.21	173.97	55.3	83.66:Ne 4.24:Ha 1.83:Hb2 1.83:Hb3 1.61:Hba 1.84:Hbb 1.58:Hg2 1.58:Hg3 1.37:Hga 1.56:Hgb 3.02:Hda 3.16:Hdb 7.79:He 28.97:Cb 25.22:Cg 42.38:Cd
200	**His**	7.46	118.65	174.02	52.92	4.85:Ha 2.55:Hba 3.20:Hbb 7.37:Hd2 7.97:He1 27.47:Cb 118.45:Cd2 137.15:Ce1
201	**Gly**	8.14	107.16	172.39	44.71	3.83:Haa 4.07:Hab
202	**Gly**	7.85	106.51	172.53	43.1	3.41:Haa 4.52:Hab
203	**Ser**	7.54	112.03	172.31	54.63	4.81:Ha 3.71:Hba 4.08:Hbb 61.60:Cb
204	**Ile**	10.83	132	176.08	62.52	4.14:Ha 1.58:Hb 1.11:Hg12 1.11:Hg13 0.58:Hg2* 0.60:Hd1* 37.16:Cb 28.50:Cg1 14.33:Cg2 12.84:Cd1
205	**Ile**	7.89	125.38	175.39	64.31	3.30:Ha 1.60:Hb 0.72:Hg1a 1.11:Hg1b 1.15:Hg2* 1.12:Hd1* 37.26:Cb 27.89:Cg1 15.92:Cg2 13.55:Cd1
206	**Glu**	8.05	120.5	176.86	56.77	4.11:Ha 2.06:Hba 2.20:Hbb 2.21:Hga 2.37:Hgb 27.50:Cb 33.97:Cg
207	**Ala**	7.94	119.13	178.89	53.25	4.16:Ha 1.57:Hb* 17.10:Cb
208	**Gln**	7.32	115.37	176.82	57.38	3.84:Ha 2.51:Hba 2.66:Hbb 2.21:Hga 2.66:Hgb 30.12:Cb 35.10:Cg

RESIDUE NUMBER	RESIDUE	H	N	C	CA	OTHERS
209	**Val**	8.65	123.59	176.05	56.28	3.77:Ha 2.23:Hb 0.97:Hga* 1.19:Hgb* 30.33:Cb 19.21:Cga 20.51:Cgb
210	**Gln**	8.78	117.19	176.97	56.46	111.93:Ne2 3.99:Ha 2.06:Hb2 2.06:Hb3 2.48:Hga 2.56:Hgb 6.81:He2a 7.69:He2b 25.95:Cb 32.16:Cg
211	**Arg**	7.22	118.63	176.15	57.14	3.92:Ha 1.32:Hba 1.74:Hbb 1.41:Hga 2.06:Hgb 26.46:Cb 29.81:Cg
212	**Ala**	7.59	120.38	179.28	53.1	4.05:Ha 1.54:Hb* 16.35:Cb
213	**Val**	8.44	118.2	175.23	63.55	3.75:Ha 2.10:Hb 1.01:Hg1* 1.01:Hg2* 0.96:Hga* 1.06:Hgb* 29.89:Cb 19.14:Cga 20.36:Cgb
214	**Ala**	7.33	119.82	174.89	50.08	4.33:Ha 1.42:Hb* 17.37:Cb
215	**Asp**	7.98	120.09	172.98	52.82	4.24:Ha 2.41:Hba 3.10:Hbb 37.89:Cb
216	**Ala**	7.78	120.22	-	48.41	4.61:Ha 1.35:Hb* 16.81:Cb
217	**Pro**	-	-	175.57	60.5	4.55:Ha 1.85:Hba 2.40:Hbb 2.06:Hg2 2.06:Hg3 3.63:Hda 3.85:Hdb 30.07:Cb 25.57:Cg 48.57:Cd
218	**Trp**	8.9	121.91	175.24	56.45	129.42:Ne1 4.09:Ha 3.10:Hb2 3.12:Hb3 7.43:Hd1 10.36:He1 7.31:He3 7.65:Hz2 6.92:Hz3 7.52:Hh2 26.75:Cb 125.09:Cd1 117.14:Ce3 113.09:Cz2 118.45:Cz3 122.94:Ch2
219	**Asp**	8.72	125.95	174.78	52.16	4.71:Ha 2.77:Hba 2.92:Hbb 41.08:Cb
220	**Glu**	9.1	124.85	176.93	58.38	3.98:Ha 2.12:Hb2 2.12:Hb3 2.39:Hg2 2.39:Hg3 27.85:Cb 34.45:Cg
221	**Gln**	8.55	119.23	176.01	56.29	115.27:Ne2 4.18:Ha 2.31:Hb2 2.31:Hb3 2.50:Hg2 2.50:Hg3 7.04:He2a 7.84:He2b 26.89:Cb 32.14:Cg
222	**Leu**	8.17	119.99	175.97	56.62	4.00:Ha 1.66:Hba 2.19:Hbb 1.79:Hg 0.37:Hda* 0.74:Hdb* 39.36:Cb 25.26:Cg 20.53:Cda 24.18:Cdb
223	**Leu**	8.43	117.68	175.84	55.89	3.99:Ha 1.46:Hba 1.76:Hbb 1.37:Hg 0.04:Hda* 0.58:Hdb* 39.56:Cb 24.65:Cg 19.64:Cda 23.42:Cdb
224	**Ala**	7.86	119.12	178.66	53.03	4.01:Ha 1.48:Hb* 16.05:Cb
225	**Asp**	8.24	118.74	176.17	55.33	4.56:Ha 2.53:Hba 2.95:Hbb 38.38:Cb
226	**Val**	8.73	118.53	177.3	65.31	3.62:Ha 2.34:Hb 1.10:Hga* 1.16:Hgb* 30.02:Cb 21.56:Cga 20.33:Cgb
227	**Ala**	8.52	120.67	175.64	52.41	4.09:Ha 1.44:Hb* 16.79:Cb
228	**Ala**	7.43	117.92	176.54	50.25	4.34:Ha 1.59:Hb* 17.50:Cb

RESIDUE NUMBER	RESIDUE	H	N	C	CA	OTHERS
229	**Val**	6.99	119.32	173.06	62.95	3.65:Ha 1.95:Hb 1.06:Hga* 1.27:Hgb* 30.14:Cb 21.78:Cg1 21.78:Cg2 21.57:Cga 22.01:Cgb
230	**Arg**	9.27	128.24	175.85	51.77	84.75:Ne 4.84:Ha 1.70:Hba 2.19:Hbb 1.91:Hg2 1.91:Hg3 3.27:Hda 3.35:Hdb 7.43:He 31.97:Cb 24.60:Cg 41.62:Cd
231	**Thr**	9.4	114.59	175.06	64.62	3.98:Ha 4.17:Hb 1.35:Hg2* 66.08:Cb 21.82:Cg2
232	**Asp**	8	114.64	174.17	53.66	4.45:Ha 2.36:Hba 2.80:Hbb 38.66:Cb
233	**Asp**	7.62	116.03	172.26	54.06	4.53:Ha 2.45:Hba 2.60:Hbb 39.33:Cb
234	**Phe**	7.69	116.29	172.81	55.63	4.94:Ha 3.04:Hb2 3.05:Hb3 6.50:Hd* 7.19:He* 7.43:Hz 39.84:Cb 127.40:Cd* 130.10:Ce* 130.10:Cz
235	**Asp**	9.19	121.04	172.43	51.06	5.06:Ha 2.40:Hba 2.77:Hbb 41.21:Cb
236	**Leu**	8.38	123.17	174.1	52.35	5.42:Ha 1.06:Hba 1.95:Hbb 1.40:Hg 0.69:Hda* 0.75:Hdb* 41.61:Cb 26.06:Cg 23.29:Cda 24.89:Cdb
237	**Thr**	9.5	123.79	-	58.4	5.07:Ha 3.99:Hb 1.31:Hg2* 70.12:Cb 21.91:Cg2
238	**Pro**	-	-	174.94	61.7	4.84:Ha 2.19:Hba 3.06:Hbb 2.02:Hga 2.20:Hgb 3.84:Hda 4.14:Hdb 31.65:Cb 26.04:Cg 50.70:Cd
239	**Ala**	8.95	126.94	177.09	52.37	4.62:Ha 1.52:Hb* 16.80:Cb
240	**Ala**	8.69	115.8	176.63	51.84	4.20:Ha 1.49:Hb* 17.16:Cb
241	**Glu**	8.01	114.42	173.93	54.51	4.31:Ha 1.86:Hba 2.19:Hbb 2.22:Hga 2.38:Hgb 29.93:Cb 34.29:Cg
242	**Arg**	7.54	120.25	171.55	54.3	85.99:Ne 4.22:Ha 1.24:Hba 1.50:Hbb 1.05:Hg2 1.05:Hg3 2.44:Hd2 2.44:Hd3 6.91:He 31.41:Cb 24.48:Cg 41.37:Cd
243	**Asp**	8.1	122.71	176.14	53.28	4.59:Ha 2.62:Hba 2.94:Hbb 38.56:Cb
244	**Ile**	9.27	131.62	175.97	60.29	4.33:Ha 1.91:Hb 1.20:Hg1a 1.43:Hg1b 0.91:Hg2* 0.83:Hd1* 37.35:Cb 27.79:Cg1 15.31:Cg2 12.74:Cd1
245	**Asp**	8.72	123.22	176.43	55.26	4.54:Ha 2.76:Hba 2.89:Hbb 37.78:Cb
246	**Asp**	7.93	119.55	173.54	52.29	4.77:Ha 3.11:Hb2 3.11:Hb3 39.53:Cb
247	**Ser**	7.69	114.98	172.73	57.49	2.82:Ha 3.26:Hba 3.43:Hbb 63.53:Cb
248	**Leu**	8.89	118.27	176.07	54.22	4.61:Ha 1.57:Hba 1.83:Hbb 1.75:Hg 0.83:Hda* 0.92:Hdb* 39.64:Cb 24.72:Cg 22.26:Cda 24.27:Cdb

RESIDUE NUMBER	RESIDUE	H	N	C	CA	OTHERS
249	**Asp**	6.66	114.09	170.56	51.11	5.18:Ha 2.07:Hb2 2.07:Hb3 42.43:Cb
250	**Leu**	9.01	119.59	174.64	51.45	4.59:Ha 0.93:Hba 1.67:Hbb 1.36:Hg 0.58:Hda* 0.76:Hdb* 43.55:Cb 24.67:Cg 23.30:Cda 21.92:Cdb
251	**Val**	9.06	121.12	172.03	58.51	4.72:Ha 1.75:Hb 0.51:Hga* 0.57:Hgb* 32.67:Cb 19.49:Cga 19.19:Cgb
252	**Ala**	9.32	126.51	172.2	48.3	5.13:Ha 1.02:Hb* 20.03:Cb
253	**Ile**	8.1	118.63	172.45	58.59	4.49:Ha 1.66:Hb 0.79:Hg1a 1.35:Hg1b 0.71:Hg2* 0.77:Hd1* 37.50:Cb 24.69:Cg1 15.22:Cg2 12.43:Cd1
254	**Ala**	9.35	130.56	172.4	47.6	5.41:Ha 1.16:Hb* 20.60:Cb
255	**Val**	9.5	120.76	174.1	58.61	5.20:Ha 2.10:Hb 0.97:Hg1* 0.97:Hg2* 0.91:Hga* 1.00:Hgb* 33.66:Cb 18.85:Cga 19.70:Cgb
256	**Arg**	9.01	130.77	-	51.16	86.60:Ne 4.57:Ha 1.18:Hba 1.74:Hbb 0.86:Hga 1.45:Hgb 2.02:Hda 2.44:Hdb 6.34:He 28.97:Cb 24.90:Cg 41.10:Cd

Appendix-II Expression and Purification of Ulp1 Protease

This appendix describes the expression and purification protocol of the Ulp1 protease that cleaves the SUMO tag. The protease was used to cleave the tag (wherever mentioned) during purification of SUMO-tagged proteins described in this study.

Expression and purification protocol

Culture growth: A loop of frozen BL21 cells of the expression clone were plated on LB-agar kanamycin plate and incubated overnight at 37°C. All the cells were collected from the plate and inoculated in 100 ml LB media with 100 µg/ml kanamycin. The culture was grown overnight at 37°C, with constant shaking at 220 rpm. The culture was scaled up to 2 litres by adding 30 ml of the overnight culture in each 1-liter fresh LB media with kanamycin. Cells were grown at 37°C till OD_{600} reached 0.8, cooled down to 20°C and, induced with 0.5 mM IPTG for protein expression, and grown overnight at 20°C.

Cell harvesting and Lysis: Cells were harvested by centrifugation (4200 x g, 30 min, 4°C) and the pellet was suspended in buffer (50 mM Tris-HCl pH 7.5, 20 mM Imidazole, 500 mM NaCl, cocktail of protease inhibitor), 10 µg/mL DNaseI and 5 mM $MgCl_2$ were added to dissolved pellet. The cells were lysed by passing them through Emulsiflex four times, followed by centrifugation (20,000 x g, 25 min, at 4°C).

Affinity purification: Pre-packed sepharose column (HiTrap IMAC FF, 5ml, GE Healthcare) was charged with 0.1M NiSO4, according to the manufacturer's protocol, and equilibrated. Supernatant was bound to the column with help of a peristaltic pump at room temperature. Column was washed for 30 CV with wash buffer (50 mM Tris-HCl pH 7.5, 20 mM Imidazole, 500 mM NaCl), and 15 ml protein was eluted with elution buffer (50mM Tris-HCl pH 7.5, 500 mM Imidazole, 500 mM NaCl). The purity of the protein was determined by SDS-PAGE. Protein concentration was measured at 280 nm with NanoDrop spectrophotometer.

Buffer Exchange: Protein's buffer was exchanged with the buffer containing 10% glycerol, 75 mM Tris-HCl pH 8.0, 0.5 mM DTT and 1 mM EDTA by centrifugation (5000 x g, 30 min, 4°C) in 10 kDa cutoff concentrators (Vivaspin, 20 ml, Satorius stedim). 12mg of protein was purified.

Storage of protein- The protein was stored in the buffer containing 50% glycerol, 75 mM Tris-HCl pH 8.0, 0.5 mM DTT and 1 mM EDTA. Protein was aliquoted in 100 µl fractions, snap froze in liquid N_2 and stored in -80°C.

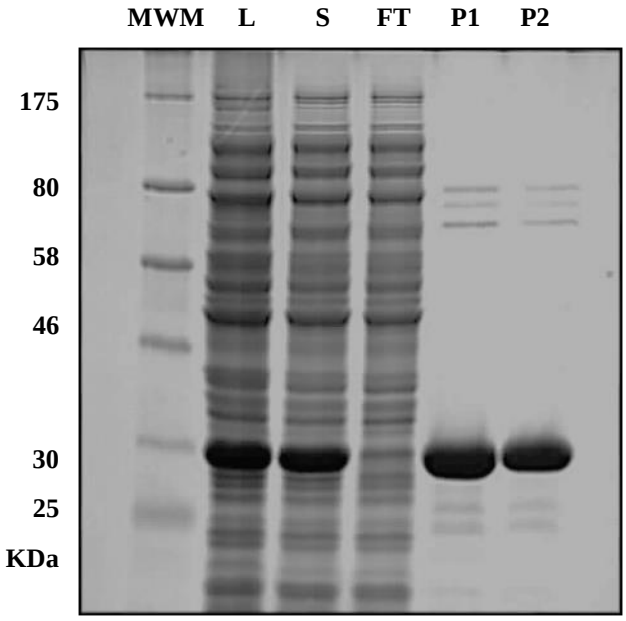


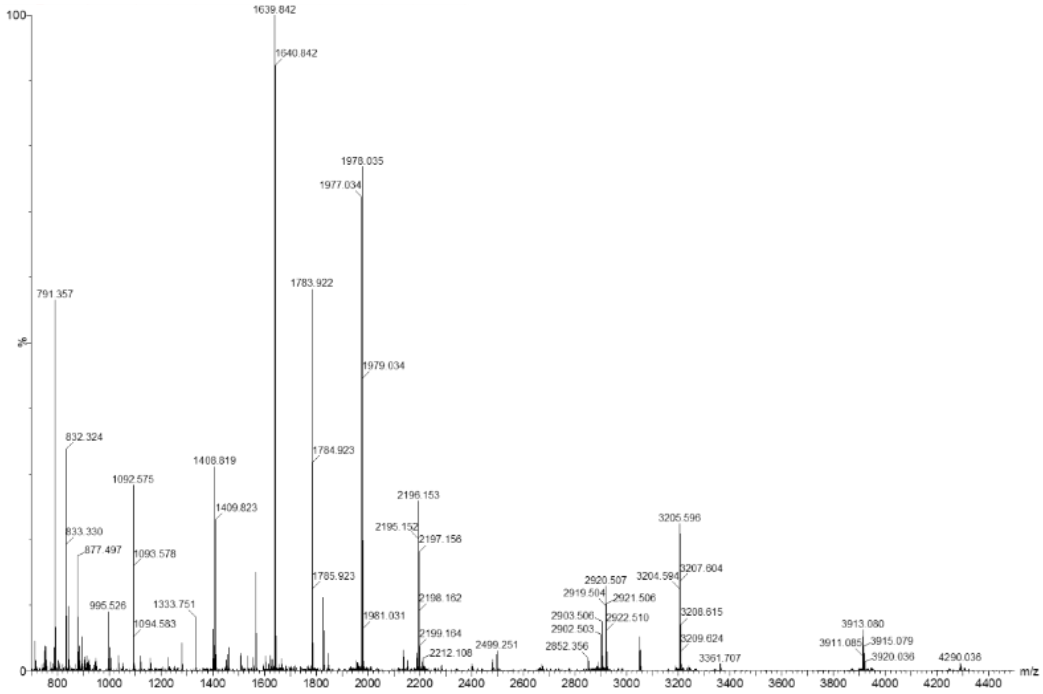
Figure AII Expression and purification of Ulp1 protease.

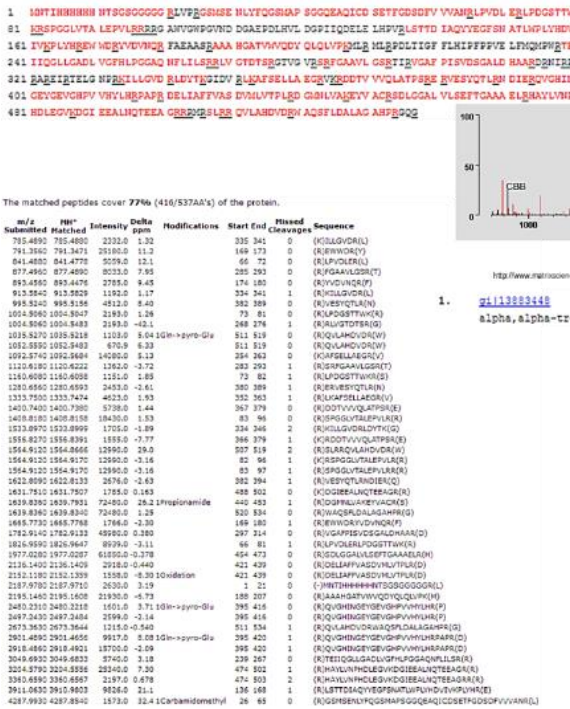
Metal affinity purification of His_6 tagged protein on Ni-sepharose column. (MWM: molecular weight marker, L: lysate, FT: flow through, S: supernatant, P1 and P2: protein)

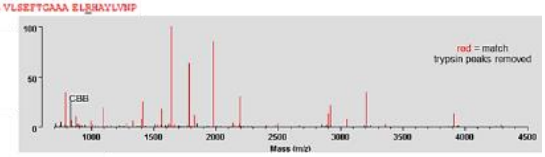
Appendix-III Protein Analysis Data of OtsA and Rv3030

This appendix reports the supporting information on the MALDI fingerprinting and intact mass analysis of OtsA and Rv3030.

1. MALDI fingerprinting of *M. tuberculosis* OtsA (full-length, pHAT-4 construct) **(Chapter-2, Section 2.5.2)**

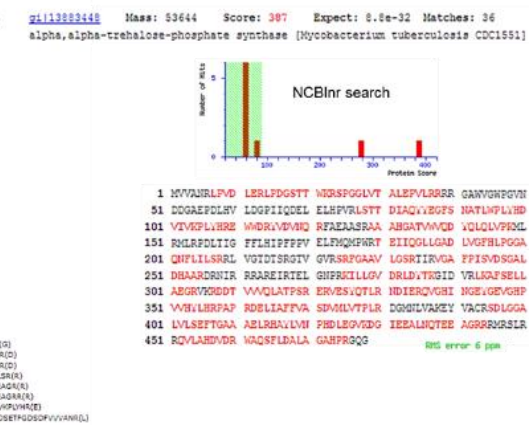




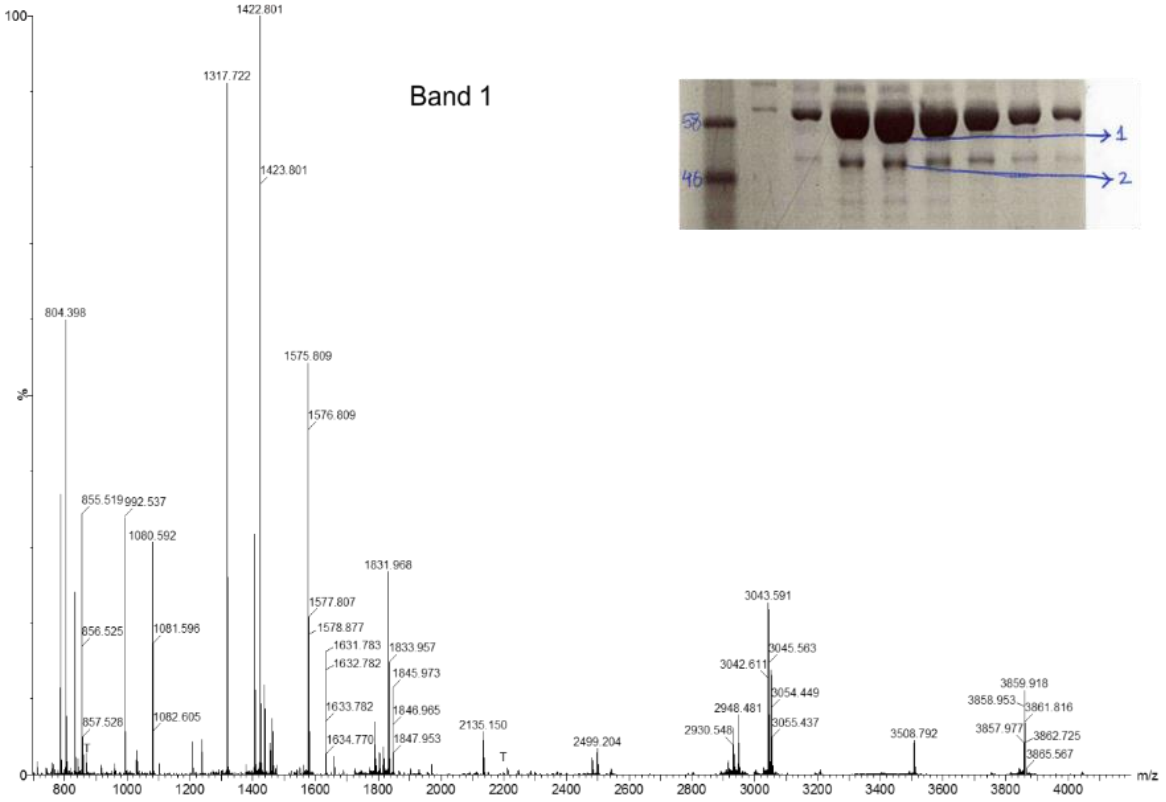


The matched peptides cover **77%** (416/537AA's) of the protein.

m/z Submitted	MH+ Matched	Intensity	Delta ppm	Modifications	Start	End	Missed Cleavages	Sequence
795.4890	795.4890	2232.0	1.32		335	341	0	(K)DLGVDR(L)
791.3560	791.3471	25180.0	11.2		169	173	0	(R)EWWDR(Y)
841.4880	841.4778	5059.0	12.1		66	72	0	(R)LPVDLER(L)
877.4960	877.4890	8033.0	7.95		285	293	0	(R)FGAAVLGSR(T)
893.4560	893.4476	2785.0	9.45		174	180	0	(R)VVDVMIQR(F)
913.5840	913.5829	1192.0	1.17		334	341	1	(R)KDLGVDR(L)
995.5240	995.5156	4512.0	8.40		382	389	0	(R)VESYQTLR(N)
1004.9060	1004.5047	2193.0	1.26		73	81	0	(K)LPDGSTTWK(R)
1004.9060	1004.5483	2193.0	-42.1		268	276	1	(R)ALVSTDTER(G)
1035.5270	1035.5218	1103.0	5.04 1Gln->pyro-Glu		511	519	0	(R)QVLAMDVDR(W)
1052.5592	1052.5483	670.9	6.33		511	519	0	(R)QVLAMDVDR(W)
1092.5742	1092.5684	14080.0	5.13		254	262	0	(K)AFSELLAEGR(V)
1120.6180	1120.6222	1262.0	-3.72		283	293	1	(R)SRFGAAVLGSR(T)
1160.6080	1160.6098	1151.0	1.85		73	82	1	(K)LPDGSTTWKR(S)
1280.6560	1280.6593	2453.0	-2.61		380	389	1	(R)ERVESYQTLR(N)
1333.7500	1333.7474	4623.0	1.93		352	363	1	(R)LKAFSELLAEGR(V)
1403.7400	1403.7360	5738.0	1.44		367	379	0	(K)DDTVVVQLATPSR(E)
1408.8180	1408.8158	18430.0	1.53		83	96	0	(R)SPGGLVTALEPVLR(R)
1533.0970	1533.8999	1705.0	-1.89		334	346	2	(R)KDLGVDRLDYTK(G)
1556.8270	1556.8391	1555.0	-7.77		366	379	1	(K)KDDTVVVQLATPSR(E)
1564.9120	1564.8666	12590.0	29.0		507	519	2	(R)SLRRQVLAHDVDR(W)
1564.9120	1564.9170	12590.0	-3.16		82	96	1	(K)RSPGGLVTALEPVLR(R)
1564.9120	1564.9170	12590.0	-3.16		83	97	1	(R)SPGGLVTALEPVLRR(R)
1622.8090	1622.8133	2676.0	-2.63		382	394	1	(R)VESYQTLRNDIER(Q)
1631.7510	1631.7507	1755.0	0.163		488	502	0	(K)DGIERALNQTEAGR(R)
1639.8360	1639.7931	72480.0	26.2 1Propionamide		440	453	1	(R)DGPNVLVAKEYVAQR(S)
1639.8360	1639.8340	72480.0	1.25		520	534	0	(R)WAQSFLDALAGAHPR(G)
1665.7730	1665.7768	1766.0	-2.30		169	180	1	(R)EWWDRVVDVMIQ(R)
1782.9140	1782.9133	49980.0	0.280		297	314	0	(R)VGAPPSEVDSGALDHAAR(D)
1826.9590	1826.9647	8929.0	-3.11		66	81	1	(K)LPVDLERLPDGSTTWK(R)
1977.0280	1977.0287	61850.0	-0.378		454	473	0	(R)SDLGGALVLSEFTGAAELR(H)
2136.1400	2136.1409	2918.0	-0.440		421	439	0	(R)DELAFFVASDVMLVTPLR(D)
2152.1180	2152.1359	1558.0	-8.30 1Oxidation		421	439	0	(R)DELAFFVASDVMLVTPLR(D)
2187.9780	2187.9710	2630.0	3.19		1	21	0	(-)MNTIHHHHHHNTSGSGGGGGR(L)
2195.1460	2195.1608	21930.0	-5.73		188	207	0	(R)AAAHGATVWVQDYQLQLVPK(M)
2480.2310	2480.2218	1601.0	3.71 1Gln->pyro-Glu		395	416	0	(R)QVGHINGEYGEVGHPVVHYLHR(P)
2497.2430	2497.2484	2599.0	-2.14		395	416	0	(R)QVGHINGEYGEVGHPVVHYLHR(P)
2673.3630	2673.3644	1215.0	-0.540		511	534	1	(R)QVLAMDVDRWAQSFLDALAGAHPR(G)
2931.4690	2931.4656	9917.0	5.08 1Gln->pyro-Glu		395	420	1	(R)QVGHINGEYGEVGHPVVHYLHRPAPR(D)
2918.4860	2918.4921	15700.0	-2.09		395	420	1	(R)QVGHINGEYGEVGHPVVHYLHRPAPR(D)
3048.6930	3048.6833	5740.0	3.18		239	267	0	(K)TEEIQGLLGADLVGFHLPGGAQNFLILSRR(K)
3204.5790	3204.5836	29340.0	7.30		474	502	1	(R)HAVLVNPHDLEGVKDGIEEALNQTEEAGRR(R)
3360.6590	3360.6567	2197.0	0.678		474	503	2	(R)HAVLVNPHDLEGVKDGIEEALNQTEEAGRRR(R)
3911.0630	3910.9803	9826.0	21.1		136	168	1	(R)LSTTDSAQYVEGFSNATLNPLYHDVTVHPLYHR(E)
4287.9930	4287.8940	1573.0	32.4 1Carbamidomethyl		26	65	0	(K)GSHSENLYFQGSHAPSGGQEAQICDSETFGDSDFVVVAMR(L)

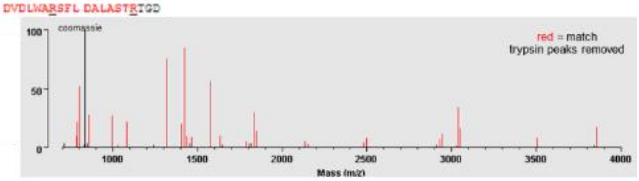


2. MALDI fingerprinting of *M. thermoresistibile* OtsA (full-length, untagged, pET-28a SUMO construct) **(Chapter-2, Section 2.5.3)**



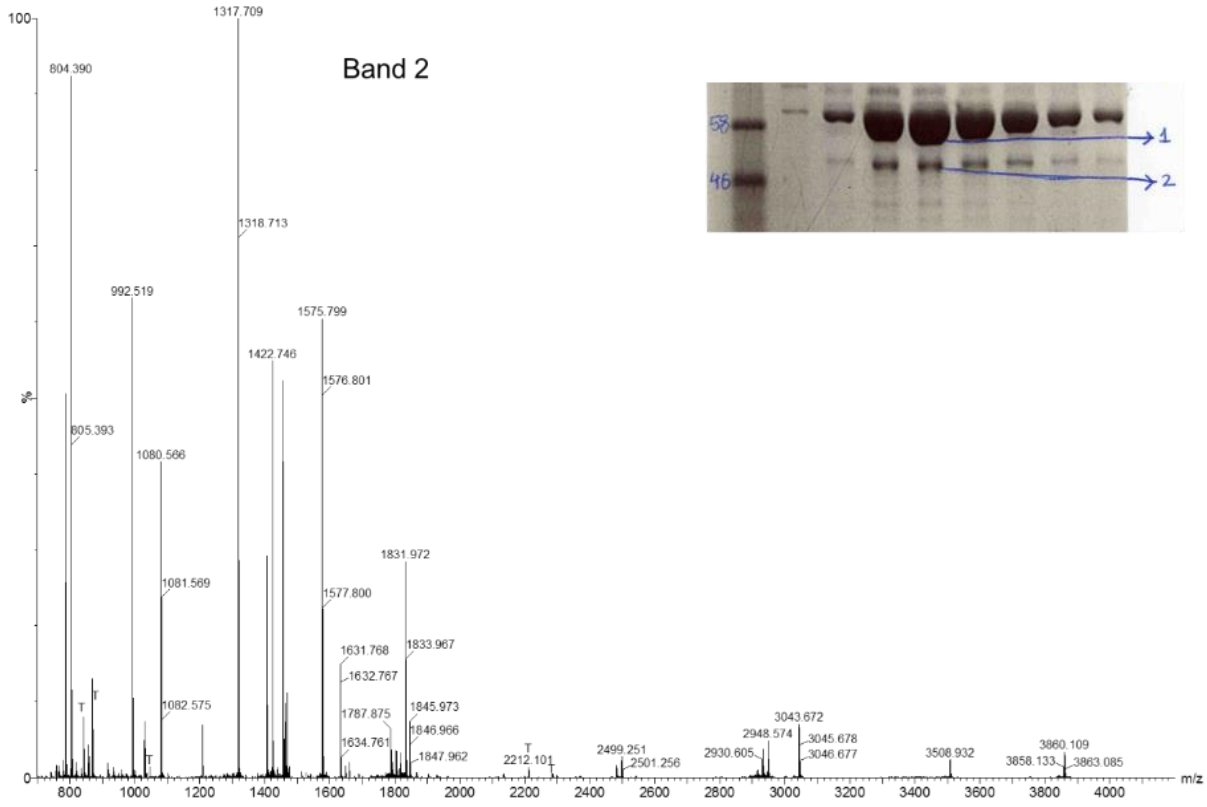
1 MADRGDSGDS DFVVVANRLP IDLERAPDGT TSWKRSPGGL VTALEPLLRR RRGAWIGWPG IPDSDEDPIV DGDLVLYPVR
81 LSADDVAQVV EGFSNATLWP LYHDVIVKPI YNRQWWERYV EVNRRFAEAT SRAAARGATV WVQDYQLQLV PEMLRELRPD
161 LTIGFFLHIP FPPVELFMQL PWRTEITDGL LGADLVGFHL PGGAQNFLFL ARRLVGANTS RASVGVRSKF GEVQIGSRTV
241 KVGAFPISID SADLDRQARQ RSIRQRARQI RAELGNPRRI LLGVDRLDYT KGIDVRLQAF AELLAEGKVN REDTVFVQLA
321 TPSRERVEAY RLLRDDIERQ VGHINGEYGE VGHPVVHYLH RPVPRRELIA FFVAADVMLV TPLRDGMNLV AKEYVACRSD
401 LGGALVLSEF TGAAAELGQA YLVNPHNLDH VKDTIVAALN QTPEEGRRRM RALRRQVLAR DVDLWARSFL DALASTRTGD
481 ADAVPV

Band 1



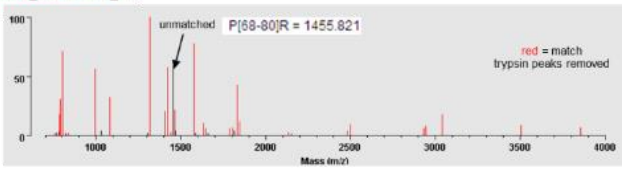
The matched peptides cover **70%** (341/486AA's) of the protein.

m/z Submitted	MH+ Matched	Intensity	Delta ppm	Modifications	Start	End	Missed Cleavages	Sequence
785.4940	785.4880	6056.0	7.68		280	286	0	(R)ILLGVDR(L)
787.3510	787.3522	13950.0	-1.52	1Gln->pyro-Glu	114	118	0	(R)QWWER(Y)
804.3710	804.3787	32540.0	-9.63		114	118	0	(R)QWWER(Y)
855.4950	855.4934	17390.0	1.82		19	25	0	(R)LPIDLER(A)
992.5140	992.5160	17140.0	-1.99		230	238	0	(K)FGEVQIGSR(T)
1029.5740	1029.5687	1554.0	5.11		332	339	1	(R)LLRDDIER(Q)
1080.5710	1080.5684	13720.0	2.41		468	477	0	(R)SFLDALASTR(T)
1317.7130	1317.7161	47420.0	-2.37		297	308	0	(R)LQAFAELLAEGR(V)
1405.7200	1405.7223	12610.0	-1.62	1Gln->pyro-Glu	456	467	0	(R)QVLAHDVDLWAR(S)
1422.7860	1422.7488	53410.0	26.1		456	467	0	(R)QVLAHDVDLWAR(S)
1437.6660	1437.6605	6168.0	3.85		5	18	0	(R)GDSGDSDFVVVANR(L)
1462.7510	1462.7536	5434.0	-1.80		312	324	0	(R)EDTVFVQLATPSR(E)
1575.7980	1575.8013	35550.0	-2.10		242	256	0	(K)VGAFPISIDSADLDR(Q)
1631.7660	1631.7694	6554.0	-2.06		433	447	0	(K)DTMVAALNQTPEEGR(R)
1647.7480	1647.7643	1329.0	-9.88	1Oxidation	433	447	0	(K)DTMVAALNQTPEEGR(R)
1787.8630	1787.8705	3453.0	-4.18		433	448	1	(K)DTMVAALNQTPEEGRR(R)
1803.8730	1803.8654	2101.0	4.22	1Oxidation	433	448	1	(K)DTMVAALNQTPEEGRR(R)
1831.9470	1831.9661	18570.0	-10.4		309	324	1	(R)VNREDTVFVQLATPSR(E)
1844.9590	1844.9905	8749.0	-17.1		137	152	0	(R)GATVWVQDYQLQLVPK(M)
2134.1620	2134.1617	3322.0	0.152		366	384	0	(R)EELIAFFVAADVMLVTPLR(D)
2150.1370	2150.1566	1649.0	-9.11	1Oxidation	366	384	0	(R)EELIAFFVAADVMLVTPLR(D)
2480.2040	2480.2218	2632.0	-7.18	1Gln->pyro-Glu	340	361	0	(R)QVGHINGEYGEVGHPVVHYLHR(P)
2497.2040	2497.2484	5089.0	-17.8		340	361	0	(R)QVGHINGEYGEVGHPVVHYLHR(P)
2929.5290	2929.4969	4875.0	11.0	1Gln->pyro-Glu	340	365	1	(R)QVGHINGEYGEVGHPVVHYLHRPVPR(E)
2946.4670	2946.5234	7306.0	-19.1		340	365	1	(R)QVGHINGEYGEVGHPVVHYLHRPVPR(E)
3026.6930	3026.6001	1200.0	30.7	1Oxidation	159	183	0	(R)PDLTIGFFLHIPFPPVELFMQLPWR(T)
3042.5900	3042.5950	21790.0	-1.65	2Oxidation	159	183	0	(R)PDLTIGFFLHIPFPPVELFMQLPWR(T)
3042.5900	3042.6047	21790.0	-4.84		184	212	0	(R)TEITDGLLGADLVGFHLPGGAQNFLFLAR(R)
3051.4750	3051.5098	10290.0	-11.4		53	80	0	(R)GAWIGWPGIPDSDEDPIVDGDLVLYPVR(L)
3506.7460	3506.7914	5094.0	-13.0		399	432	0	(R)SDLGGALVLSEFTGAAAELGQAYLVNPHNLDHVK(D)
3857.8770	3857.9174	10770.0	-10.5		81	113	1	(R)LSADDVAQYYEGFSNATLWPLYHDVIVKPIYNR(Q)




```
1   MADRGDSGDS DFVVVANRLP IDLERAPDGT TSWKRSPGGL VTALEPLLRR RRGAWIGWPG IPDSDEDPIV DGDLVLVPVR
81  LSADDVAQYY EGFSNATLWP LYHDVIVKPI YNRQWWERYV EVNRRFAEAT SRAAARGATV WVQDYQLQLV PKMLRELRPD
161 LTIGFFLHIP FPPVELFMQL PWRTEITDGL LGADLVGFHL PGGAQNFLFL ARRLVGANTS RASVGVRSKF GEVQIGSRTV
241 KVGAFPISID SADLDRQARQ RSIRQRARQI RAELGNPRRI LLGVDRLDYT KGIDVRLQAF AELLAEGRVN REDTVFVQLA
321 TPSRERVEAY RLLRDDIERQ VGHINGEYGE VGHPVVHYLH RPVPREELIA FFVAADVMLV TPLRDGMNLV AKEYVACRSD
401 LGGALVLSEF TGAAAELGQA YLVNPHNLDH VKDTMVAALM QTPEEGRRRM RALRRQVLAH DVDLWARSFL DALASTRTGD
481 ADAVPV
```

Band 2

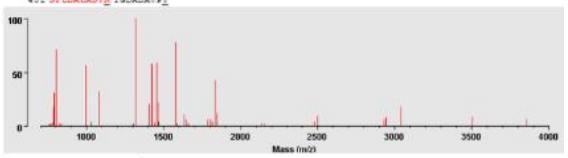


The matched peptides cover **63%** (305/486AA's) of the protein.

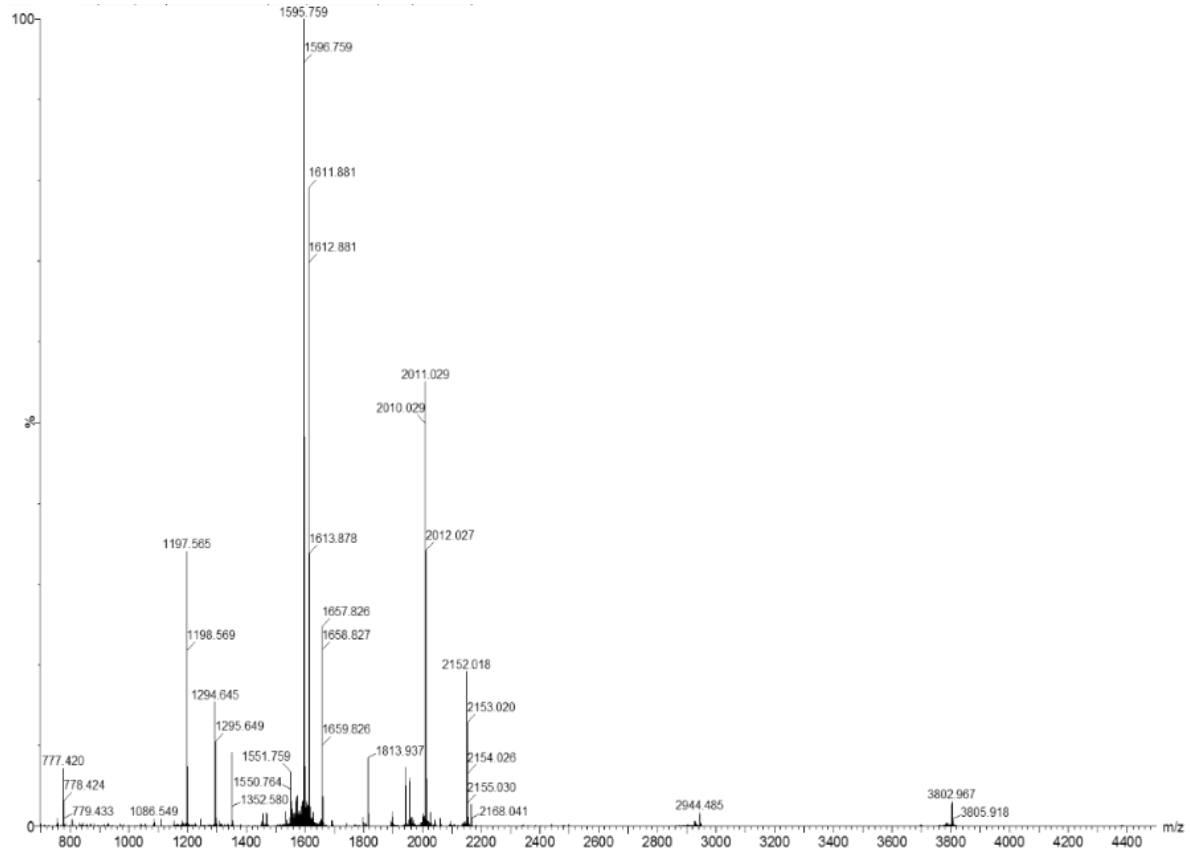
m/z Submitted	MH+ Matched	Intensity	Delta ppm	Modifications	Start	End	Missed Cleavages	Sequence
756.4020	756.3999	2666.0	2.81		272	278	0	(R)AELGNPR(R)
779.4080	779.4046	3980.0	4.33		119	124	0	(R)YVEVNR(R)
785.4950	785.4880	20600.0	8.96		280	286	0	(R)ILLGVDR(L)
787.3610	787.3522	35090.0	11.2	1Gln->pyro-Glu	114	118	0	(R)QWWER(Y)
804.3840	804.3787	81730.0	6.53		114	118	0	(R)QWWER(Y)
820.3740	820.3737	2842.0	0.410	1Oxidation	114	118	0	(R)QWWER(Y)
836.3760	836.3686	2575.0	8.87	2Oxidation	114	118	0	(R)QWWER(Y)
992.5230	992.5160	64250.0	7.08		230	238	0	(K)FGEVQIGSR(T)
1029.5700	1029.5687	3054.0	1.23		332	339	1	(R)LLRDDIER(Q)
1080.5770	1080.5684	36540.0	7.96		468	477	0	(R)SFLDALASTR(T)
1317.7280	1317.7161	113800.0	9.01		297	308	0	(R)LQAFAELLAEGR(V)
1405.7390	1405.7223	23460.0	11.9	1Gln->pyro-Glu	456	467	0	(R)QVLAHDVDLWAR(S)
1422.7620	1422.7488	65590.0	9.26		456	467	0	(R)QVLAHDVDLWAR(S)
1438.7450	1438.7437	3030.0	0.877	1Oxidation	456	467	0	(R)QVLAHDVDLWAR(S)
1462.7630	1462.7536	24730.0	6.40		312	324	0	(R)EDTVFVQLATPSR(E)
1575.8130	1575.8013	89030.0	7.42		242	256	0	(K)VGAFPISIDSADLDR(Q)
1631.7790	1631.7694	12520.0	5.91		433	447	0	(K)DTMVAALNQTPEEGR(R)
1647.7570	1647.7643	7197.0	-4.42	1Oxidation	433	447	0	(K)DTMVAALNQTPEEGR(R)
1787.8950	1787.8705	6914.0	13.7		433	448	1	(K)DTMVAALNQTPEEGRR(R)
1803.8890	1803.8654	7483.0	13.1	1Oxidation	433	448	1	(K)DTMVAALNQTPEEGRR(R)
1831.9760	1831.9661	48800.0	5.41		309	324	1	(R)VNREDTVFVQLATPSR(E)
1844.9730	1844.9905	13990.0	-9.49		137	152	0	(R)GATVWVQDYQLQLVPK(M)
2134.1840	2134.1617	3290.0	10.5		366	384	0	(R)EELIAFFVAADVMLVTPLR(D)
2150.1530	2150.1566	2267.0	-1.67	1Oxidation	366	384	0	(R)EELIAFFVAADVMLVTPLR(D)
2480.2470	2480.2218	4597.0	10.2	1Gln->pyro-Glu	340	361	0	(R)QVGHINGEYGEVGHPVVHYLHR(P)
2497.2510	2497.2484	10740.0	1.06		340	361	0	(R)QVGHINGEYGEVGHPVVHYLHR(P)
2929.5860	2929.4969	6696.0	30.4	1Gln->pyro-Glu	340	365	1	(R)QVGHINGEYGEVGHPVVHYLHRPVPR(E)
2946.5240	2946.5234	9114.0	0.203		340	365	1	(R)QVGHINGEYGEVGHPVVHYLHRPVPR(E)
3042.6190	3042.5950	20130.0	7.88	2Oxidation	159	183	0	(R)PDLTIGFFLHIPFPPVELFMQLPWR(T)
3042.6190	3042.6047	20130.0	4.69		184	212	0	(R)TEITDGLLGADLVGFHLPGGAQNFLFLAR(R)
3506.7840	3506.7914	10310.0	-2.12		399	432	0	(R)SDLGGALVLSEFTGAAAELGQAYLVNPHNLDHVK(D)
3857.9220	3857.9174	7931.0	1.20		81	113	1	(R)LSADDVAQYYEGFSNATLWPLYHDVIVKPIYNR(Q)

Better match to truncated protein 68-486

```
1   PIVDGDLVLV PVRLSADDVA QVYEGFSNAT LWPLYHDVIV KPIYNRQWWE RVVEVNRRFA EATSRAAARG ATVWVQDYQL
81  QLVPKMLREL RPDLTIGFFL HIPFPPVELF MQLPWRTEIT DGLLGADLVG FHLPGGAQNF LFLARRLVGA NTSRASVGVR
161 SKFGEVQIGS RTVKVGAFPI SIDSADLDRQ ARQRSIRQRA RQIRAELGNP RRILLGVDRL DYTKGIDVRL QAFAELLAEG
241 RVNREDTVFV QLATPSRERV EAVRLLRDDI ERQVGHINGE YGEVGHPVVH YLHRPVPREE LIAFFVAADV MLVTPLRDGM
321 NLVAKEYVAC RSDLGGALVL SEFTGAAAEL GQAYLVNPHN LDHVKDTMVA ALMQTPEEGR RRMRALRRQV LAHDVDLWAR
401 SFLDALASTR TGDADAVPV
```



3. MALDI fingerprinting of Rv3030 orthologue in *M. smegmatis* (unlabelled, untagged, full-length pET-28a SUMO construct) **(Chapter-3, Section 3.5.4.1)**

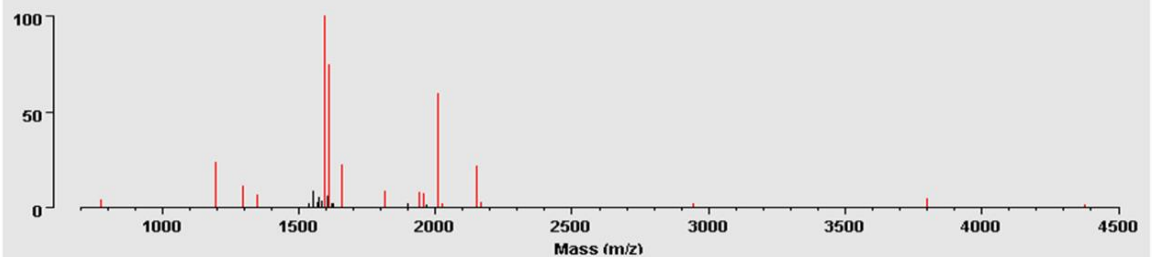


Name: **M. smegmatis MC2-155 | MSMEG_2350**

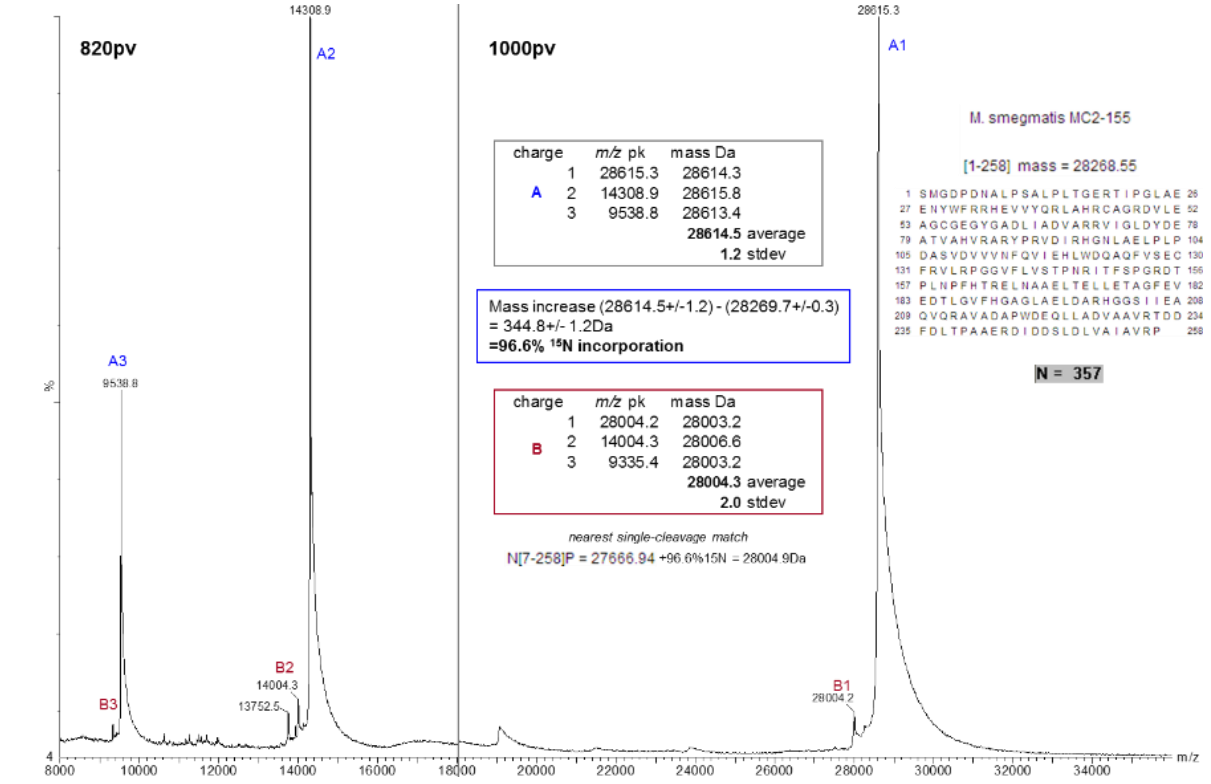
1 SMGDPDNALP SALPLTGERT IPGLAEENYW FRRHEVVYQR LAHRCAGRDV LEAGCGEGYG ADLIADVARR VIGLDYDEAT
81 VAHVRARYPR VDIRHGNLAE LPLPDASVDV VVNFQVIEHL WDQAQFVSEC FRVLRPGGVF LVSTPNRITF SPGRDTPLNP
161 FHTRELNAAE LTELLETAGF EVEDTLGVFH GAGLAELDAR HGGSIIEAQV QRAVADAPWD EQLLADVAAV RTDDFDLTPA
241 AERDIDDSLD LVAIAVRP

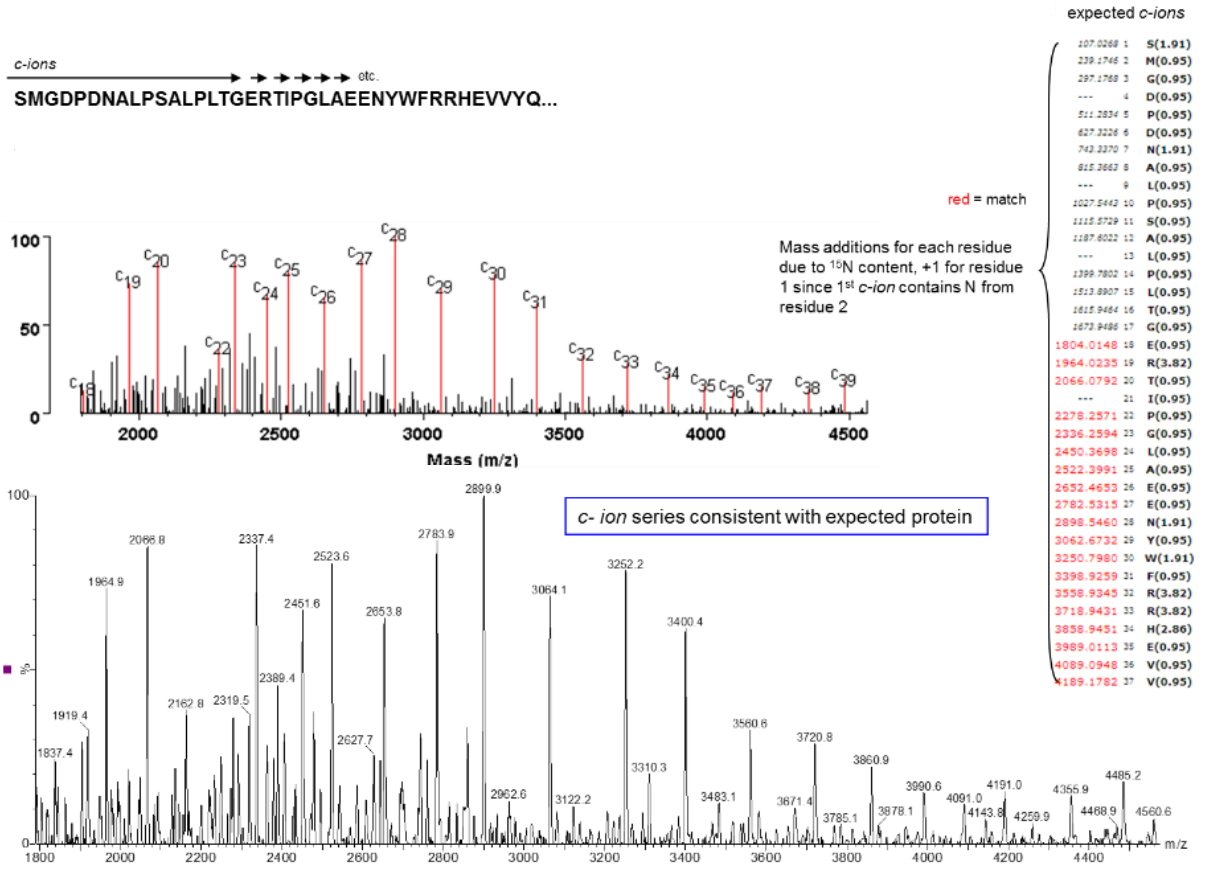
The matched peptides cover **90%** (233/258AA's) of the protein.

m/z Submitted	MH+ Matched	Intensity	Delta ppm	Modifications	Start	End	Missed Cleavages	Sequence
777.4200	777.4254	13230.0	-6.90		148	154	0	(R)ITFSPGR(D)
1197.5640	1197.6011	74490.0	-31.0		155	164	0	(R)DTPLNPFHTR(E)
1294.6430	1294.6862	36550.0	-33.4		201	212	0	(R)HGGSIIEAQVQR(A)
1350.5690	1350.6172	21970.0	-35.7		232	243	0	(R)TDDFDLTPAAER(D)
1595.7540	1595.7853	315800.0	-19.6		20	32	0	(R)TIPGLAEENYWFR(R)
1611.8760	1611.8588	236800.0	10.7		244	258	0	(R)DIDDSLDLVAIAVRP(-)
1611.8760	1611.9329	236800.0	-35.3		133	147	0	(R)VLRPGGVFLVSTPNR(I)
1657.8220	1657.8544	71110.0	-19.6		71	85	0	(R)VIGLDYDEATVAHVR(A)
1813.9300	1813.9555	27030.0	-14.1		70	85	1	(R)RVIGLDYDEATVAHVR(A)
1940.9180	1940.9382	26760.0	-10.4		1	19	0	(-)SMGDPDNALPSALPLTGER(T)
1956.9130	1956.9331	23660.0	-10.3	1Oxidation	1	19	0	(-)SMGDPDNALPSALPLTGER(T)
2010.0210	2010.0291	187900.0	-4.02		213	231	0	(R)AVADAPWDEQLLADVAAVR(T)
2026.0050	2026.0240	6951.0	-9.38	1Oxidation	213	231	0	(R)AVADAPWDEQLLADVAAVR(T)
2151.0100	2151.0023	69000.0	3.59	1Carbamidomethyl	49	69	0	(R)DVLEAGCGEGYGADLIADVAR(R)
2165.0220	2165.0179	10240.0	1.88	1Propionamide	49	69	0	(R)DVLEAGCGEGYGADLIADVAR(R)
2943.4740	2943.4582	7312.0	5.37		232	258	1	(R)TDDFDLTPAAERDIDDSLDLVAIAVRP(-)
3800.9510	3800.8865	14910.0	17.0		165	200	0	(R)ELNAAELTELLETAGFEVEDTLGVFHGAGLAELDAR(H)
4379.2600	4379.1554	5918.0	23.9	1Carbamidomethyl	95	132	0	(R)HGNLAELPLPDASVDVVVNFQVIEHLWDQAQFVSECFR(V)

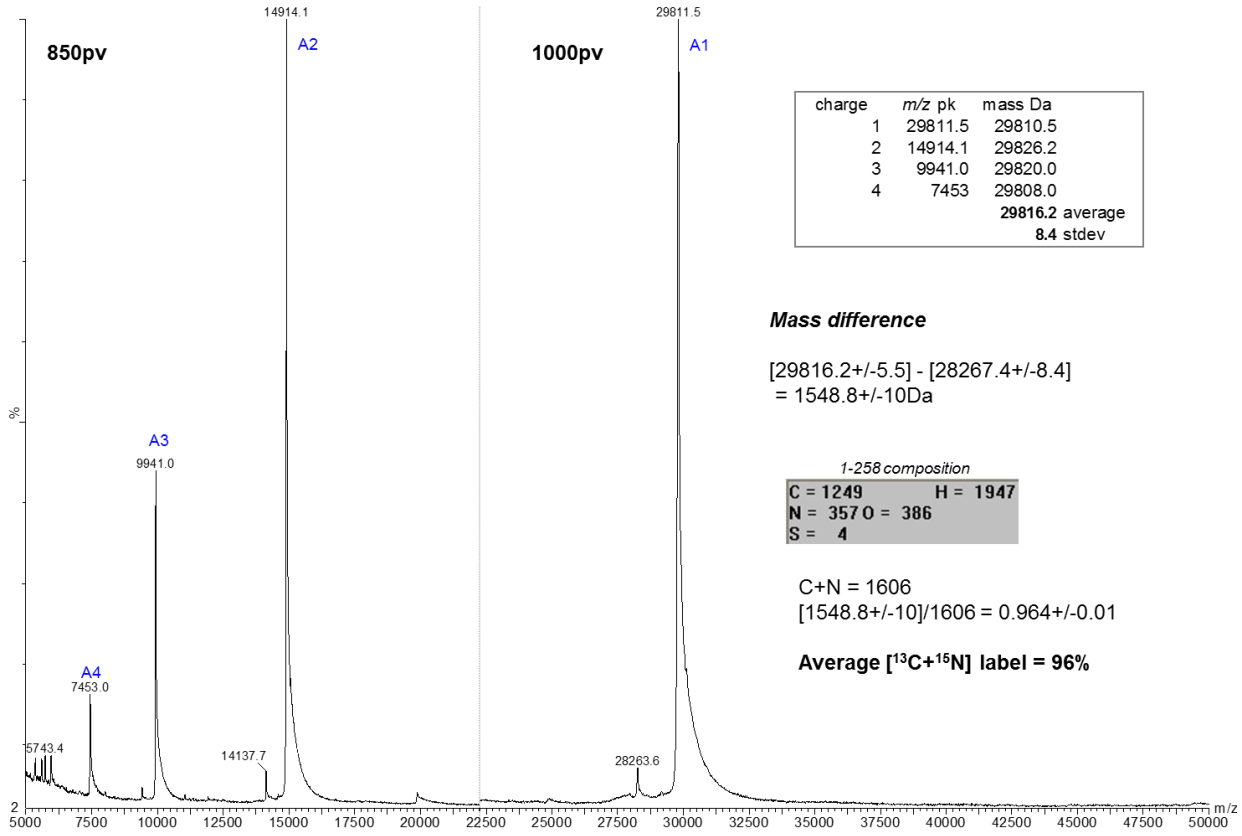


4. Intact mass analysis of ^{15}N-labelled, full-length orthologue of Rv3030 in *M. smegmatis* **(Chapter-4, Section 4.5.2)**





5. Intact mass analysis of ^{15}N, ^{13}C-labelled, full-length orthologue of Rv3030 in *M. smegmatis* **(Chapter-4, Section 4.5.2)**



Appendix-IV Composition of Commonly Used Reagents

The culture media and solutions used during this study are described in this appendix. The media were autoclaved and solutions were filter sterilized before use. Filtered milliQ water was used to make the media and solutions.

Luria-Bertani (LB) Medium

10 g/L	Tryptone
5 g/L	Yeast extract
10 g/L	NaCl

The pH was adjusted to 7.2 and autoclaved.

Luria-Bertani Agar

10 g/L	Tryptone
5 g/L	Yeast extract
10 g/L	NaCl
20 g/L	Agar

The pH was adjusted to 7.2 and autoclaved.

20 X M9 Salts

120 g/L	Na_2HPO_4
60 g/L	KH_2PO_4
10 g/L	NaCl

The solution was autoclaved.

100 X Micronutrients Solution (200 ml)

160 mg	$FeCl_2$
324 mg	$MnCl_2$
10 mg	$ZnSO_4$
3 mg	$CoCl_2$
2 mg	H_3BO_3

Solution was filtered sterilized, and stored at 4°C.

1000 X Vitamins Solution (50 ml)

50 mg Thiamine Pyrophosphate

50 mg Niacinamide

35 mg Biotin

Solution was filtered sterilized, and stored at 4°C.

Antibiotic Solutions

1. Ampicillin: 100 mg/ml stock solution was prepared in sterile milliQ water and filter sterilized. Stored in aliquots at -20°C.
2. Kanamycin: 30 mg/ml stock solution was prepared in sterile milliQ water and filter sterilized. Stored in aliquots at -20°C.

Agarose Gel

1% agarose gel was prepared by dissolving 1 g agarose in 100 ml 1 X TBE buffer. 7.5 µl ethidium bromide was added for every 100 ml of the gel.

SDS-PAGE Gel

RESOLVING GEL (12%)

	per gel
Water	2.47 ml
Acrylamide	3 ml
3M Tris-HCl pH 8.8	1.87 ml
10% SDS	150 µl
25% APS	60 µl
TEMED	15 µl

STACKING GEL

	per gel
Water	1 ml
Acrylamide	310 µl
0.5M Tris-HCl pH 6.8	450 µl
10% SDS	19.25 µl
25% APS	8 µl
TEMED	2.25 µl

1. 30% Acrylamide Mix (Severn Biotech Ltd)
Stored at 4°C.

2. 25% Ammonium persulphate (APS) (Sigma)

2.5 g of APS was dissolved in 10 ml milliQ water. Solution was filtered sterilized, and stored in aliquots at -20°

3. 10% Sodium Dodecyl Sulphate (SDS) (Fisher)
5 g of SDS was dissolved in 50 ml milliQ water and was filtered sterilized.

4. 3 M Tris-HCl pH 8.8 (for resolving gel) (Melford)
36.34 g of Tris-HCl was dissolved in milliQ water. pH was adjusted to 8.8 and volume was made to 100 ml with milliQ water. Solution was filtered sterilized.

5. 0.5 M Tris-HCl pH 6.8 (for stacking gel) (Melford)
6.06 g of Tris-HCl was dissolved in milliQ water. pH was adjusted to 6.8 and volume was made to 100 ml with milliQ water. Solution was filtered sterilized.

SDS-PAGE Running Buffer (1 X)

3.02 g/L	Tris-HCl Base
14.4 g/L	Glycine
1 g/L	SDS

4 X SDS-PAGE Loading Dye (containing β-mercaptoethanol)

62.5 mM	Tris-HCl.HCl pH 6.8
2%	SDS
10%	Glycerol
10%	β-mercaptoethanol
~ 0.2%	Bromophenol blue

2 X SDS-PAGE Loading Dye (containing DTT)

100 mM	Tris-HCl.HCl pH 6.8
200 mM	DTT
4%	SDS
20%	Glycerol
~ 0.2%	Bromophenol blue

1 X Coomassie Brilliant Blue-G Quick Stain

10%	Ethanol
30 mM	HCl
0.01%	Coomassie Brilliant Blue-G

Other Commonly Used Reagents

1. 1 M IPTG stock
   2.383 g IPTG was dissolved in 10 ml sterile milliQ water. Solution was filter sterilized, and stored in aliquots at -20°C.
2. 0.1 M nickel (II) sulphate hexahydrate [$NiSO_4.6H_2O$]
   1.31 g of $NiSO_4.6H_2O$ was dissolved in 50 ml sterile milliQ water. Solution was filter sterilized.

Appendix-V Targets Rv3037c and Rv3038c of MGLP Biosynthetic Pathway

AV.1 Objectives

At the outset, these projects aimed at investigation of structural properties. This appendix reports analyses of the primary structures of the protein and its orthologues in order to understand the domain architecture and residue conservation, and homology modelling of the protein structure. The design of constructs of full-length and truncated proteins is also described, followed by attempts to purify soluble and un-degraded protein.

AV.2 Methodology

AV.2.1 Bioinformatics analysis of primary structure of target proteins

The protein sequences of Rv3037c and Rv3038c were obtained from TubercuList database, and of their orthologues in *M. smegmatis (MSMEG_2330* and *MSMEG_2329*, respectively*)* from SmegmaList database, by searching with their gene names. Orthologous proteins in *M. thermoresistibile* were identified by performing a protein-BLAST (blastp) search with respective *M. tuberculosis* protein sequences; and amino acid sequences were obtained from NCBI database. Sequences were aligned using Clustal Omega. Methyltransferase family and domain alignment coordinates were identified with Pfam.

The physicochemical properties of proteins were calculated using ProtParam, and secondary structures and disordered regions of the protein were predicted using PSIpred and DisoPred, respectively.

AV.2.2 Protein Modeling

Models of full-length orthologues of Rv3037c and Rv3038c in *M. thermoresistibile* were built using ModSuite, as described in section 3.4.2.

AV.2.3 DNA, Bacterial Strains, Cloning and Expression Trials of Full-length Rv3037c and Rv3038c and their orthologues

Genomic DNA and *E. coli* strains were the same as mentioned in section 2.4.3. Dr. Michal Blaszczyk and Dr. Sachin Surade provided the purified plasmids for Rv3037c (and orthologue in *M. smegmatis*) and Rv3038c (and orthologues in *M. smegmatis* and *M. thermoresistibile*), respectively. The genes were cloned in BamHI and HindIII restriction sites in pET-28a SUMO vector.

Expression trial: E.coli BL21 (DE3) competent cells were transformed with the plasmids and cells were plated on LB agar-kanamycin plates. A protein expression trial was conducted for full-length Rv3037c, Rv3038c and their orthologues, as described in section 2.4.4.1. The identity of proteins was confirmed by MALDI fingerprinting (PNAC facility, Department of Biochemistry).

Additional expression trials were performed for Rv3037 and its orthologue in *M. smegmatis* by altering following parameters in the above-mentioned protocol, in different trials:

1. 0.1% Triton X-100 was added in the resuspension buffer.
2. Inducing the culture with 0.5 mM IPTG at 37°C for 4 hours.
3. Inducing the culture with 0.5 mM IPTG overnight (14-16 hours) at 25°C.

AV.2.4 Cloning and expression trial of full-length Rv3037c orthologue in *M. thermoresistibile*

The 1092 bp gene was PCR amplified from *M. thermoresistibile* genomic DNA with the following primer pair:
Forward Primer: 5'-TTAGGATCCATGCCGCTGACCGACGCCAC-3'
(containing BamHI restriction site- underlined)
Reverse Primer: 5'-ACTAAGCTTTCACCGTGACGGACGGCATATG-3'
(containing HindIII restriction site- underlined)

PCR was done as described in section 2.4.3, but with 15 sec annealing step at 55°C, and 40 sec extension step at 70°C. The amplicon generated was cloned in BamHI and HindIII restriction sites pET28a SUMO vector. The gene and vector were digested with the BamHI and HindIII restriction enzymes (ThermoScientific), ligated and plasmid was purified (section 2.4.3). The identity of the insert was confirmed by sequencing (DNA Sequencing Facility, Department of Biochemistry).

A protein expression trial was done as described in section 2.4.4.1, and the protein's identity was confirmed by MALDI fingerprinting (PNAC facility, Department of Biochemistry).

AV.2.5 Gene manipulation and expression trials

AV.2.5.1 N- and C- terminal domains of Rv3037c orthologue in *M. thermoresistibile*

A. N-terminal domain of Rv3037c orthologue in *M. thermoresistibile* (Residues 1-216)

The N-terminal domain was cloned in NcoI and XhoI restriction sites in pHAT5 vector. The 648 bp gene was amplified from the genomic DNA of *M. thermoresistibile* using following primer pair:
Forward primer: 5'-ATTCCATGGTGCCGCTGACCGACGCCACC-3'
(containing NcoI restriction site-underlined)

Reverse primer: 5'-ATACTCGAGGCGGCGGACGCCGGGTTCGG-3'
(containing XhoI restriction site-underlined. The stop codon was removed from the reverse primer to allow the translation of fusion protein containing C-terminal His_6-tag.)

PCR was done in the same conditions as described in section 2.4.3, but with annealing at 58°C and 20 seconds extension at 70°C. Amplicon and vector were digested with NcoI and XhoI restriction enzymes, ligated and plasmid was purified (section 2.4.3). The fidelity of the insert was confirmed by sequencing (DNA Sequencing Facility, Department of Biochemistry). A protein expression trial was done (as described in section 2.4.4.1), and protein identity was confirmed by MALDI fingerprinting (PNAC facility, Department of Biochemistry).

B. C-terminal domain of Rv3037c orthologue in *M. thermoresistibile* (Residues 217-363)

The 444 bp gene was amplified from the genomic DNA of *M. thermoresistibile* with the following primers:
Forward primer: 5'-TATGGATCCCGGGCCACCATCCTGGACCG-3'
(containing BamHI restriction site-underlined)
Reverse primer: 5'-ATCAAGCTTTCACCGTGACGGACGGCATATG-3'
(containing HindIII restriction site-underlined)

PCR was done in the same conditions as described in section 2.4.3, but with annealing at 58°C and 20 sec extension at 70°C. The amplicon generated was cloned in BamHI and HindIII restriction sites in pET28a SUMO vector. Amplified gene and vector were digested with BamHI and HindIII restriction enzymes, ligated and plasmid was purified as described in section 2.4.3. The identity of insert was confirmed by sequencing (DNA Sequencing Facility, Department of Biochemistry). A protein expression trial was carried out as detailed in section 2.4.4.1. Protein identity was confirmed by MALDI fingerprinting (PNAC facility, Department of Biochemistry).

AV.2.5.2 Truncation of Rv3038c orthologue in *M. thermoresistibile*

An 891 bp gene (111bp less from 5' end of full length 1kb gene), correlating to truncation of 37 residues from N terminal of full-length protein, was cloned in BamHI and HindIII restriction sites in the MCS of pET-28a SUMO vector. Another construct was prepared, in which 38 residues were truncated from N terminal of full-length protein, and the 888 bp gene was cloned in NcoI and XhoI restriction sites in pHAT-5 vector. Construct details and primers used are mentioned in **Table AV.1**.

Gene	Cloning vector	Residue range	Restriction sites	Primer sequence (Restriction sites underlined and bold)
Truncated 3038M.thr (37 residues truncated)	pET-28a SUMO	38-334	FP- BamHI RP- HindIII	FP:5'- ATTGGATCCCTCGCCCAGATCCTCTATCACG -3' RP:5'- ATTAAGCTTTCAGGACGGTTTGACCCCGG -3'
Truncated 3038M.thr (38 residues truncated)	pHAT-5	39-334	FP- NcoI RP- XhoI	FP: 5'- ATACCATGGCCCAGATCCTCTATCACGACTG -3' RP: 5'- ATACTCGAGGGACGGTTTGACCCCGGTG -3' (The stop codon was removed from the RP to allow the translation of a fusion protein containing a C-terminal His_6-tag)

TABLE AV.1 Truncated Rv3038c orthologue in *M. thermoresistibile* construct details and primers used
(3038M.thr refers to Rv3038c orthologue in *M. thermoresistibile*)

The gene was amplified from the genomic DNA of *M. thermoresistibile* with relevant primers, using KOD Hot Start DNA polymerase, in the same condition as described in section 2.4.3, but with annealing at 58°C and 20 sec extension. Amplicons and vectors were digested with relevant restriction enzymes, ligated and plasmid was purified as described in section 2.4.3. The identity of the insert was confirmed by sequencing (DNA Sequencing Facility, Department of Biochemistry). Protein expression trials were carried out as detailed in section 2.4.4.1. Proteins' identity was confirmed by MALDI fingerprinting (PNAC facility, Department of Biochemistry).

AV.2.6 Purification of Full-Length Recombinant Rv3038c orthologue in *M. thermoresistibile*

The recombinant orthologous protein of Rv3038c in *M. thermoresistibile*, with N-terminal, cleavable SUMO tag was purified.

Culture growth: A loop of frozen BL21 cells of the expression clone was plated on LB-agar-kanamycin plate and incubated overnight at 37°C. All the cells were collected from the plate and inoculated in 250 ml LB media with 30 μg/ml kanamycin. The culture was grown overnight at 37°C, with constant shaking at 220 rpm. The culture was scaled up to 6 litres by adding 30ml of the overnight culture in each 1-litre fresh LB media containing 30 μg/ml kanamycin. Cells were grown at 37°C till OD_{600} reached 0.63, cooled down to 20°C and, induced with 0.5 mM IPTG for protein expression, and incubated overnight (14-16 hours) at 20°C with constant shaking.

Cell harvesting and Lysis: Cells were harvested by centrifugation (4200 x g, 30 min, 4°C) during the late exponential growth phase and the pellet was re-suspended in buffer (20 mM Tris-HCl pH 7.5, 20 mM Imidazole, 500 mM NaCl, cocktail of protease inhibitor), and 10 μg/mL DNase I and 5 mM $MgCl_2$ were added to the cell suspension. The cells were lysed by passing them through Emulsiflex four times, followed by centrifugation (20,000 x g, 30 min, 4°C) to remove cell debris. Supernatant was collected and filtered using 0.2 μm syringe filter.

Affinity purification: A pre-packed sepharose column (HiTrap IMAC FF, 5 ml, GE Healthcare) was charged with 0.1 M $NiSO_4$, according to the manufacturer's protocol, and equilibrated with purification buffers. Supernatant was bound to the column with the help of a peristaltic pump at room temperature. The column was washed for 30 CV with wash buffer (50 mM Tris-HCl pH 7.5, 20 mM Imidazole, 500 mM NaCl), and 30 ml protein was eluted using elution buffer (50 mM Tris-HCl pH 7.5, 500 mM Imidazole, 500 mM NaCl). The purity of the protein was determined by SDS-PAGE. Protein concentration was measured at 280 nm with NanoDrop spectrophotometer.

Cleavage of the tag and Dialysis: Ulp1 protease was added to the eluted protein at a ratio of 1 mg Ulp1 protease to 200 mg protein to cleave the SUMO tag. The protein-protease mixture

was then dialysed in 50 mM Tris-HCl pH 7.5, 20 mM Imidazole and 500 mM NaCl buffer, using 10 kDa MWCO dialysis membrane, and incubated overnight at 4°C.

Gel Filtration: The overnight-digested protein was concentrated by centrifuging (4000 x g, 30 min, 4°C) in 10 kDa MWCO concentrators. The protein was loaded onto a Superdex-200 column, pre-equilibrated with 100 mM NaCl and 20 mM Tris-HCl pH 7.5. Purification was carried out at room temperature. The protein was eluted in 1 ml fractions and purity of fractions was determined by SDS-PAGE. The purest fractions were pooled and concentrated by centrifuging (4000 x g, 30 min, 4°C) in 10 kDa MWCO concentrators. The concentration of NaCl was reduced from 500 mM to 350 mM with 20 mM Tris-HCl pH 7.5 buffer. The purity of the protein was assessed with SDS-PAGE and final concentration was determined spectrophotometrically. The protein's identity was confirmed by MALDI fingerprinting (PNAC facility, Department of Biochemistry, University of Cambridge, UK).

In an alternative attempt of purification, only affinity purification and anionic exchange were performed. Protein was expressed and affinity purification was as described above. Then 5 mM EDTA pH 8.0 was added to protein just after elution, and anionic exchange was done.
Anionic Exchange- Protein was diluted with binding buffer (20 mM Tris-HCl pH7.5) and loaded into the column (HiTrap Q HP column, 5 ml, GE Healthcare). Column was washed with binding buffer. Protein was eluted in 5ml fractions by creating a linear gradient between the binding buffer and the elution buffer (20 mM Tris-HCl pH7.5, 1 M NaCl) over a period of 1 hour.

AV.3 Results

AV.3.1 Target: Rv3037c

AV.3.1.2 Expression trials of full-length Rv3037c and its homologs in *M. smegmatis* and *M. thermoresistibile*

An initial expression trial of SUMO-tagged *M. tuberculosis* protein and orthologues at 20°C revealed that the protein is in the inclusion bodies as no soluble expression was observed. MALDI fingerprinting confirmed that bands observed around 58 kDa were of the respective target proteins **(Figure AV.1)**.

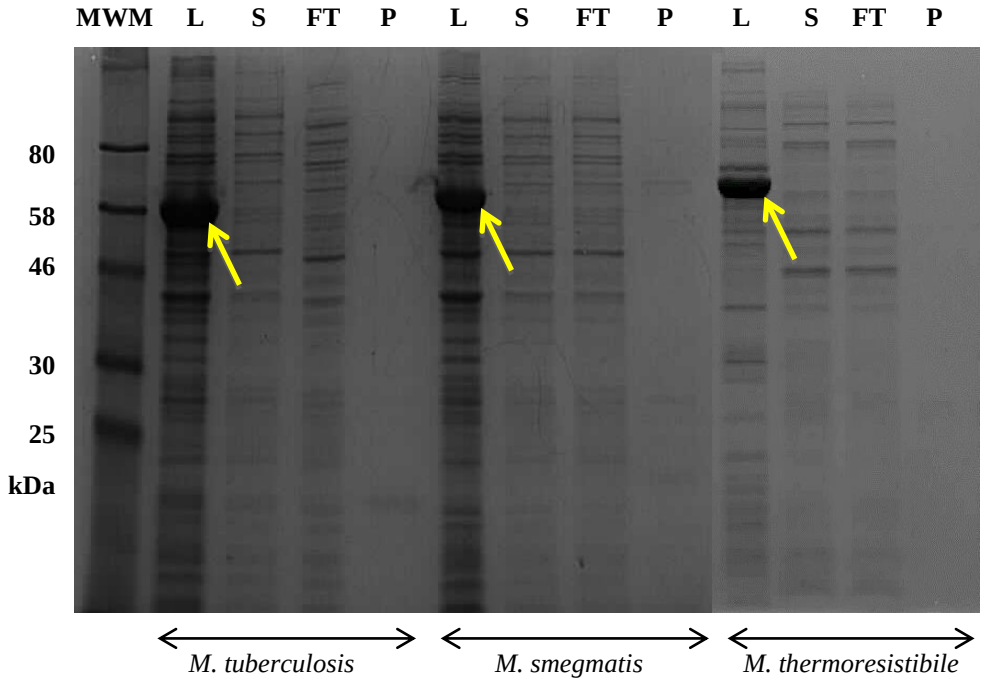
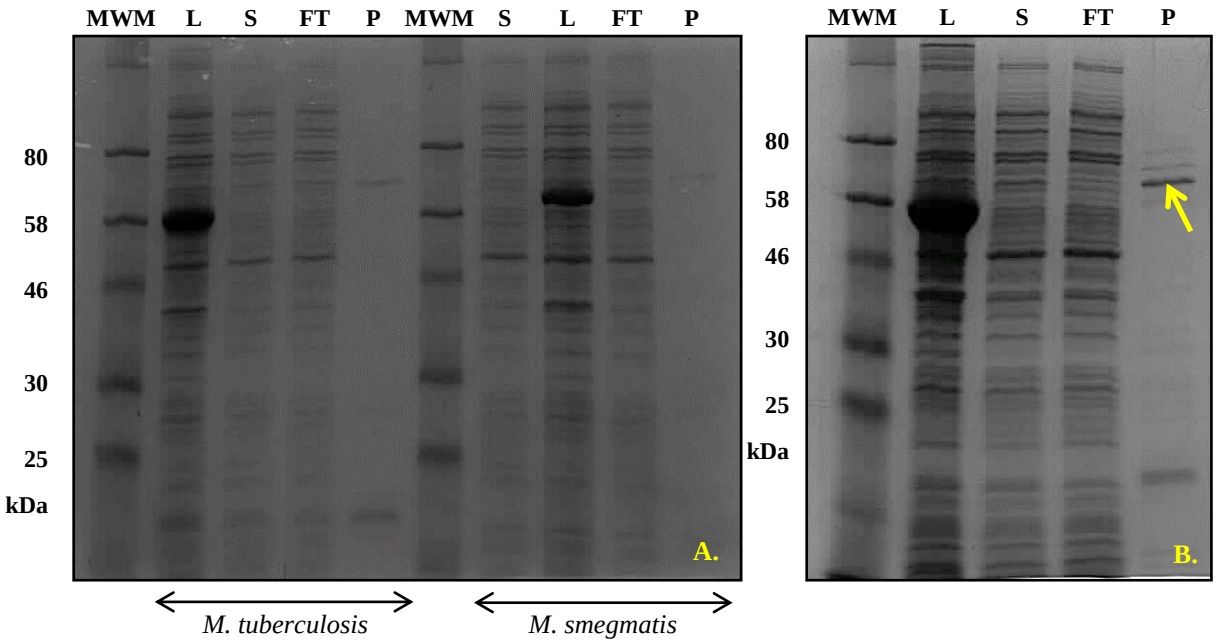


FIGURE AV.1 Expression trials of SUMO tagged Rv3037c and its orthologues.

Thick bands (yellow arrows) were seen in the respective cell lysate lanes but nothing was seen in the supernatant (soluble fraction) lanes. (MWM: molecular weight marker, L: cellular lysate, S: supernatant, FT: flow through, P: protein).

Several expression trials were carried out in attempts to express the protein in the soluble fraction; these included altering the trial conditions, such as varying the induction temperature, varying the duration for which culture was induced, and mild detergent treatment (0.1% triton X-100). Unfortunately, the attempts were not successful. A small-scale purification of *M. tuberculosis* protein was done from 1-liter culture and mass fingerprinting analysis showed that instead of the target protein, the chaperone GroEL was purified **(Figure AV.2)**.



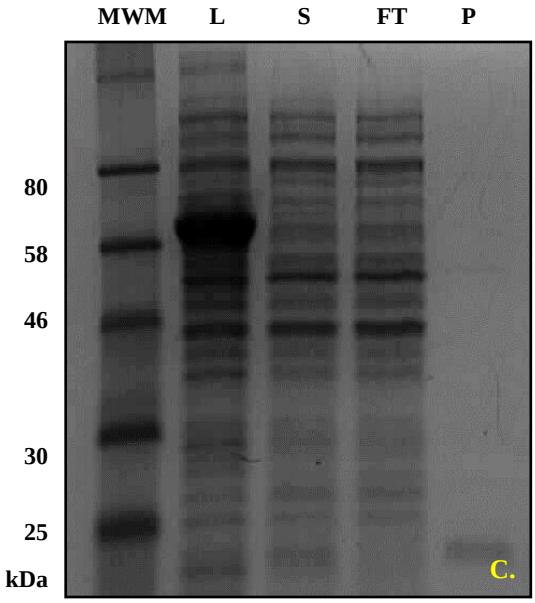


FIGURE AV.2 **Expression trials of SUMO tagged Rv3037c and its orthologue in *M. smegmatis.***

A. 0.1% Triton X-100 was added to the cell lysate to obtain soluble protein.

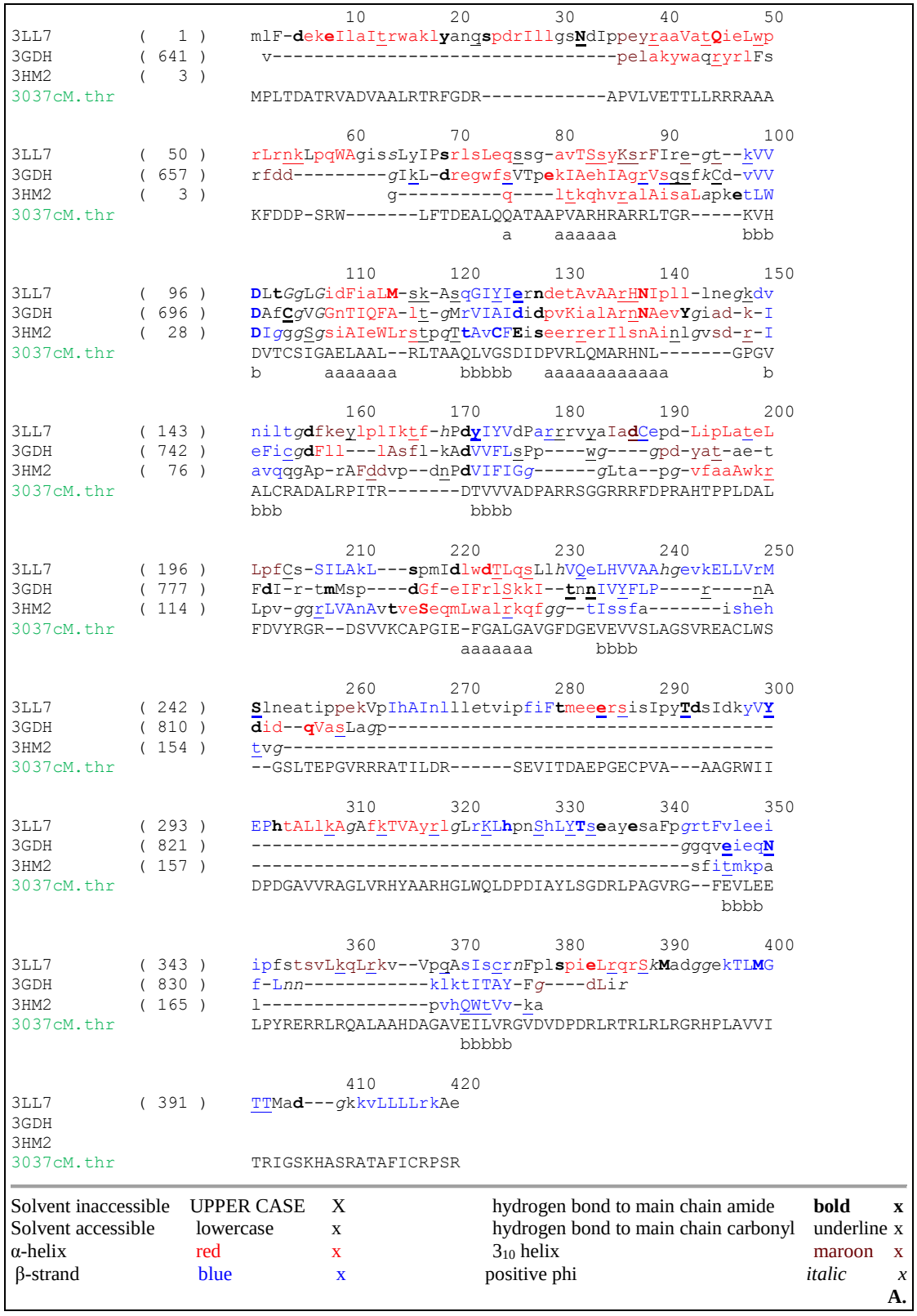
B. Protein purification of protein from 1 liter culture, induced at 37°C for 4 hours. Chaperon GroEL (marked with arrow) was obtained whereas the target protein remained insoluble.

C. Trial of Rv3037c-SUMO by inducing at 25°C overnight. Protein still remained insoluble.

(MWM: molecular weight marker, L: cellular lysate, S: supernatant, FT: flow through, P: protein).

AV.3.1.3 Protein modeling of full-length Rv3037c orthologue in *M. thermoresistibile*

The templates used for building model of target protein were methyltransferase proteins whose structures are available in PDB (PDB ID of templates- 3LL7, 3GDH, 3HM2).



Solvent inaccessible	UPPER CASE	X	hydrogen bond to main chain amide	**bold**	**x**
Solvent accessible	lowercase	x	hydrogen bond to main chain carbonyl	underline	x
α-helix	red	x	3_{10} helix	maroon	x
β-strand	blue	x	positive phi	*italic*	*x*

A.

B.

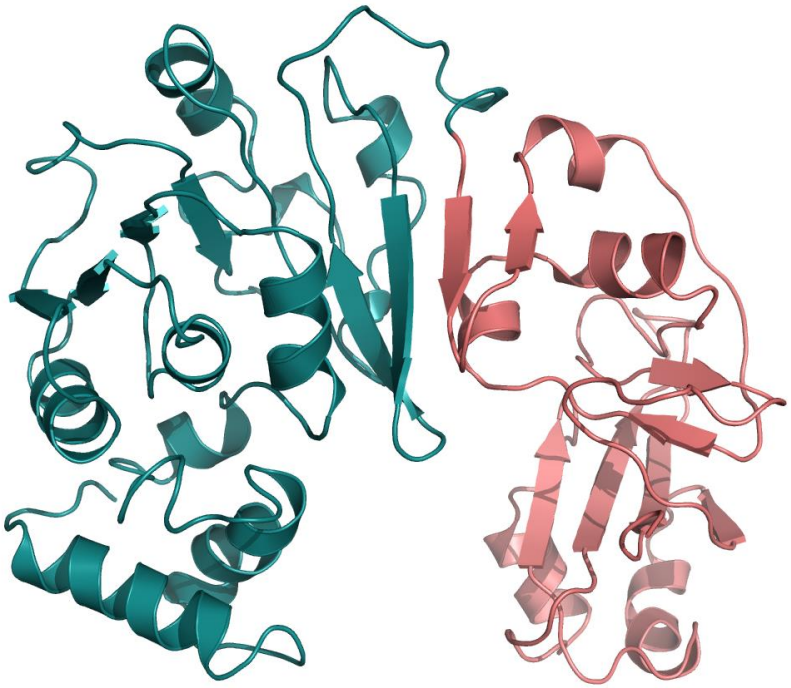


FIGURE AV.3 Model of three-dimensional structure of Rv3037c orthologue in *M. thermoresistibile.*

A. Structural alignment of target protein and templates (produced by FUGUEALI) (3037cM.thr refers to Rv3037c orthologue in *M. thermoresistibile*).

B. Model of the target protein. Probable methyltransferase domain at the N-terminus is shaded teal and the C-terminus domain is shaded pink.

PDB ID	PROTEIN	ORGANISM
3HM2_G	Putative Precorrin-6Y C5,15-Methyltransferase Targeted Domain	*Corynebacterium diphtheriae*
3GDH_C	Methyltransferase domain of human Trimethylguanosine Synthase 1 (TGS1) bound to m7GTP and adenosyl-homocysteine (active form)	*Homo sapiens*
3LL7_A	Putative methyltransferase PG_1098	*Porphyromonas gingivalis W83*

Table AV.2 Structural templates used for homology modelling of orthologue of Rv3037c in *M. thermoresistibile.*

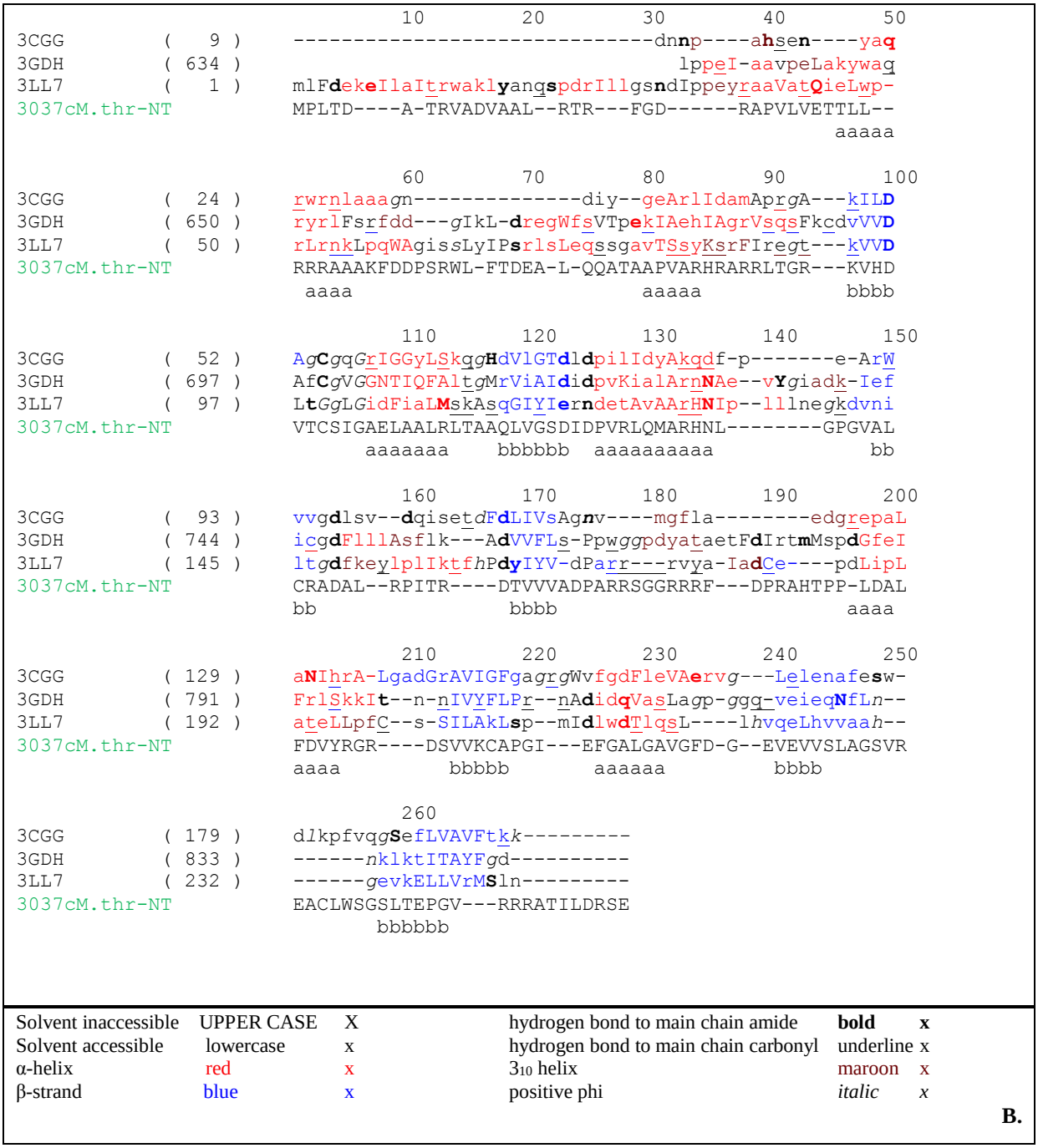
Due to low sequence similarity of Rv3037c orthologue in *M. thermoresistibile* protein with available structures (around 23% identity, determined by protein BLAST search), the model building of full-length protein was challenging. Only one template (3LL7) aligned fairly well with the entire protein, while others aligned only to the N-terminal region of the target protein. The structural alignment was not convincing, but gave a rough idea that the target protein probably has two domains: an N-terminal domain and C-terminal domain. Since the methyltransferase domain of the templates aligned with the N-terminal domain of the protein, it is probable that this is the methyltransferase domain of the protein. The C-terminal region is likely an auxiliary domain that may help in substrate binding. However, exact domain alignment coordinates are not available in the databases (search in Pfam did not yield significant matches). Thus, both the domains were individually modeled in order to get a clear overall view of the protein's three-dimensional structure.

AV.3.1.4 Protein modeling of individual domains of Rv3037c orthologue in *M. thermoresistibile*

A. Modeling of the N-terminal domain

The templates used for building model of N-terminal domain were methyltransferase proteins whose structures are available in PDB (PDB ID of templates- 3LL7, 3GDH, 3CGG). 3CGG_A is a TehB-like SAM-dependent methyltransferase (NP_600671.1) from *Corynebacterium glutamicum* ATCC 13032.

A.



Solvent inaccessible	UPPER CASE	X	hydrogen bond to main chain amide	**bold**	**x**
Solvent accessible	lowercase	x	hydrogen bond to main chain carbonyl	underline	x
α-helix	red	x	3_{10} helix	maroon	x
β-strand	blue	x	positive phi	*italic*	*x*

B.

FIGURE AV.4 Model of three-dimensional structure of N-terminal domain of Rv3037c orthologue in *M. thermoresistibile.*

A. Model of the N-terminal domain.

B. Structural alignment of target protein and templates (produced by FUGUEALI) (3037cM.thr-NT refers to N-terminus domain of Rv3037c orthologue in *M. thermoresistibile*).

B. Modeling of the C-terminal domain

No significant sequences were obtained from the BLAST search, which could serve as templates to model the C-terminal domain of the protein. Thus, only a part of a putative methyltransferase was used as a template for modeling (PDB ID- 3LL7).

```
                              10        20        30        40        50
3LL7          ( 266 )     vipfiftteersisIpyTdsidkyVYePhtaLlkagAfkTVAyrlgLrKL
3037cM.thr-CT             ---VITDAEPGECPVA---AAGRWIIDPDGAVVRAGLVRHYAARHGLWQL
                                 aaaa           bbbbb aaaaaa  aaaaaaaa  bbb

                              60        70        80        90        100
3LL7          ( 317 )     hpnShlYTseayesaFpgrtFvleeiipfstsvLkqLrkvvpq--AsIsc
3037cM.thr-CT             DPDIAYLSGDRLPAGV--RGFEVLEELPYRERRLRQALAAHDAGAVEILV
                          b  bbbbbb        bbbbbbbbbbb  333333333       bbbb

                              110       120       130       140
3LL7          ( 365 )     rnFplspieLrqrskadgge------ktLGTTadgkkvLLLlrkAe
3037cM.thr-CT             RGVDVDPDRLRTRLRLRGRHPLAVVITRIGSKHASRATAFICRPSR
                          b     aaaaaaaa            bbbbb      bbbbbbb
```

Solvent inaccessible	UPPER CASE	X	hydrogen bond to main chain amide	**bold**	**x**
Solvent accessible	lowercase	x	hydrogen bond to main chain carbonyl	underline	x
α-helix	red	x	3_{10} helix	maroon	x
β-strand	blue	x	positive phi	*italic*	*x*

A.

B.

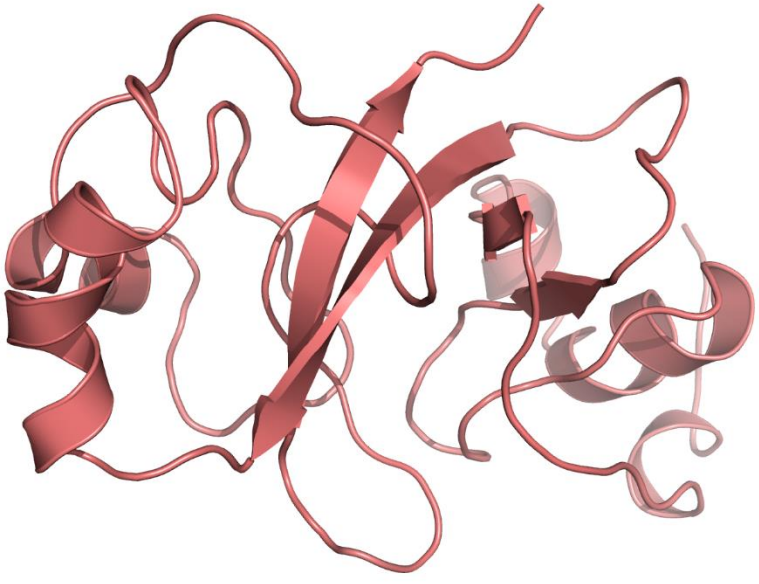


FIGURE AV.5 Model of three-dimensional structure of C-terminal domain of Rv3037c orthologue in *M. thermoresistibile*

A. Structural alignment of target protein and 3LL7 (produced by FUGUEALI) (3037cM.thr-CT refers to C-terminal domain of Rv3037c homologue in *M. thermoresistibile*).

B. Model of the C-terminal domain.

It is plausible that the full-length protein was found in inclusion bodies because the domains didn't fold properly. Thus, an attempt was made to express both domains individually in order to obtain the protein in the soluble form.

AV.3.1.5 Expression trial of N-terminal domain of Rv3037c orthologue in *M. thermoresistibile*

An expression trial of the N-terminal domain, with a C-terminal His_6 tag in *E. coli* indicated that there was no expression of the construct. The bands seen on the gel were of *E.coli* proteins, as analyzed by MALDI fingerprinting.

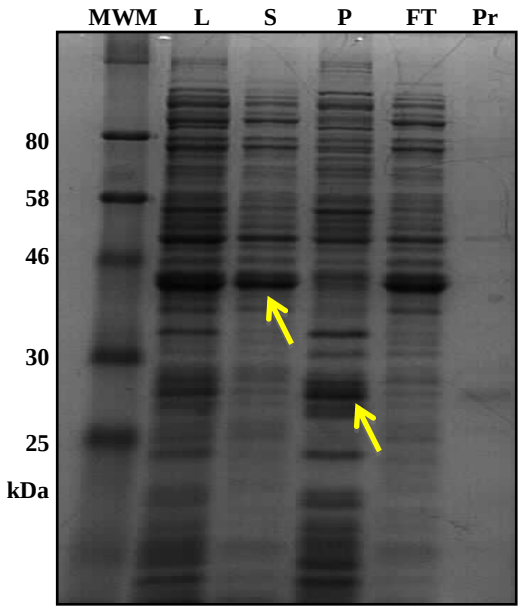


FIGURE AV.6 Expression trial of N-terminal domain of Rv3037c orthologue in *M. thermoresistibile* tagged with His_6.

Bands marked with arrows are of *E. coli* proteins.
(MWM: molecular weight marker, L: cellular lysate, S: supernatant, P: pellet, FT: flow through, Pr: protein).

AV.3.1.6 Expression trial of C-terminal domain of Rv3037c orthologue in *M. thermoresistibile*

It was possible to get expression of C-terminal domain, with a SUMO tag at its N-terminus, in the soluble fraction.

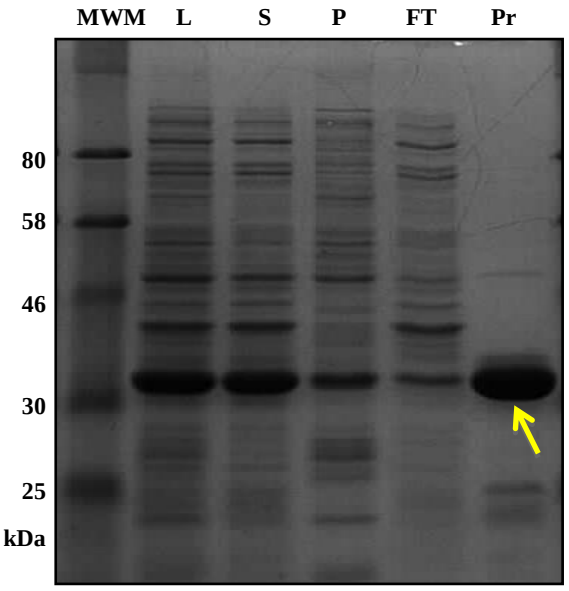


FIGURE AV.7 Expression trial of C-terminal domain of Rv3037c orthologue in *M. thermoresistibile* tagged with SUMO.
SUMO tagged C-terminal domain is seen around 30 kDa.
(MWM: molecular weight marker, L: cellular lysate, S: supernatant, P: pellet, FT: flow through, Pr: protein)

AV.3.2 TARGET: Rv3038c

AV.3.2.1 Bioinformatic Analysis

PROTEIN	METHYLTRANSFERASE DOMAIN ALIGNMENT COORDINATES (RESIDUE RANGE)
Rv3038c	82-179
Rv3038c homologue in *M. smegmatis*	81-178
Rv3038c homologue in *M. thermoresistibile*	89-186

TABLE AV.3 Alignment coordinates of the methyltransferase domain in Rv3038c and its orthologues. As given in Pfam database.

AV.3.2.2 Expression trials of full length Rv3038c and its orthologues in *M. smegmatis* and *M. thermoresistibile*

Trials for expression of the SUMO-tagged proteins indicated the *M. smegmatis* and *M. thermoresistibile* proteins expressed in the soluble fraction, but the *M. tuberculosis* protein was not identified in the soluble fraction. Bands at lower molecular weight (around 25 kDa), initially thought to be of some chaperone, were of SUMO tag along with some residues from N-terminus of the target protein that got naturally clipped when proteins were expressed in *E. coli*. This was confirmed by MALDI fingerprinting after doing large-scale expression and purification of *M. thermoresistibile* protein (discussed below).

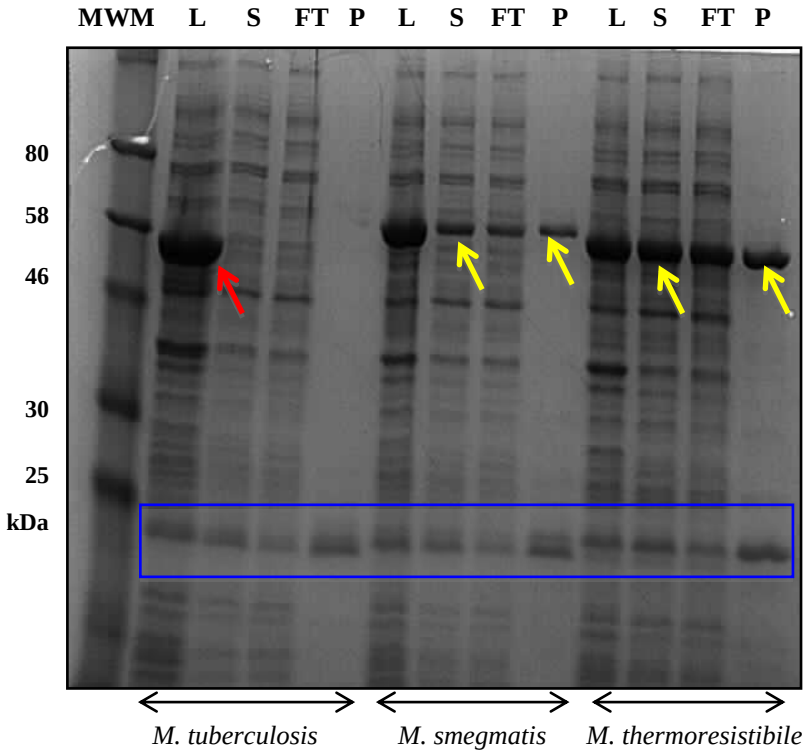


FIGURE AV.8 Expression trials of SUMO tagged Rv3038c and its orthologous proteins.

For Rv3038c, thick band (red arrow) was seen in the cell lysate lane but not in the supernatant (soluble fraction) lane. For homologous proteins, soluble expression was seen (yellow arrows). Bands at 25 kDa (blue box) were of SUMO tag with a part of the protein that got clipped at the N-terminus during expression.
(MWM: molecular weight marker, L: cellular lysate, S: supernatant, FT: flow through, P: protein).

AV.3.2.3 Expression and purification of full-length SUMO-tagged Rv3038c orthologue in *M. thermoresistibile*

High levels of expression of the orthologous *M. thermoresistibile* gene in *E. coli* were observed. However, during its expression in *E. coli*, there was partial clipping of a section at the N-terminus conjoint to SUMO tag.

The tag was cleaved off from the protein eluted by Ulp1 protease and several bands seen on the gel were analyzed by mass fingerprinting. It was inferred that there were two types of clipping of protein:

1. First clipping during expression: partial clipping of few residues from N-terminus of protein conjoint to SUMO tag during expression of full-length protein **(Figure AV.9 A)**.
2. Second clipping during Ulp1 cleavage: when SUMO-tag was cleaved from purified full-length protein by Ulp1 protease, tag got cleaved; but a few residues at the N-terminus not conjoint to the tag also got clipped **(Figure AV.9 B)**.

The gel filtration profile showed that the SUMO-tagged protein eluted just at the void volume, probably indicating its aggregation or higher molecular assembly.

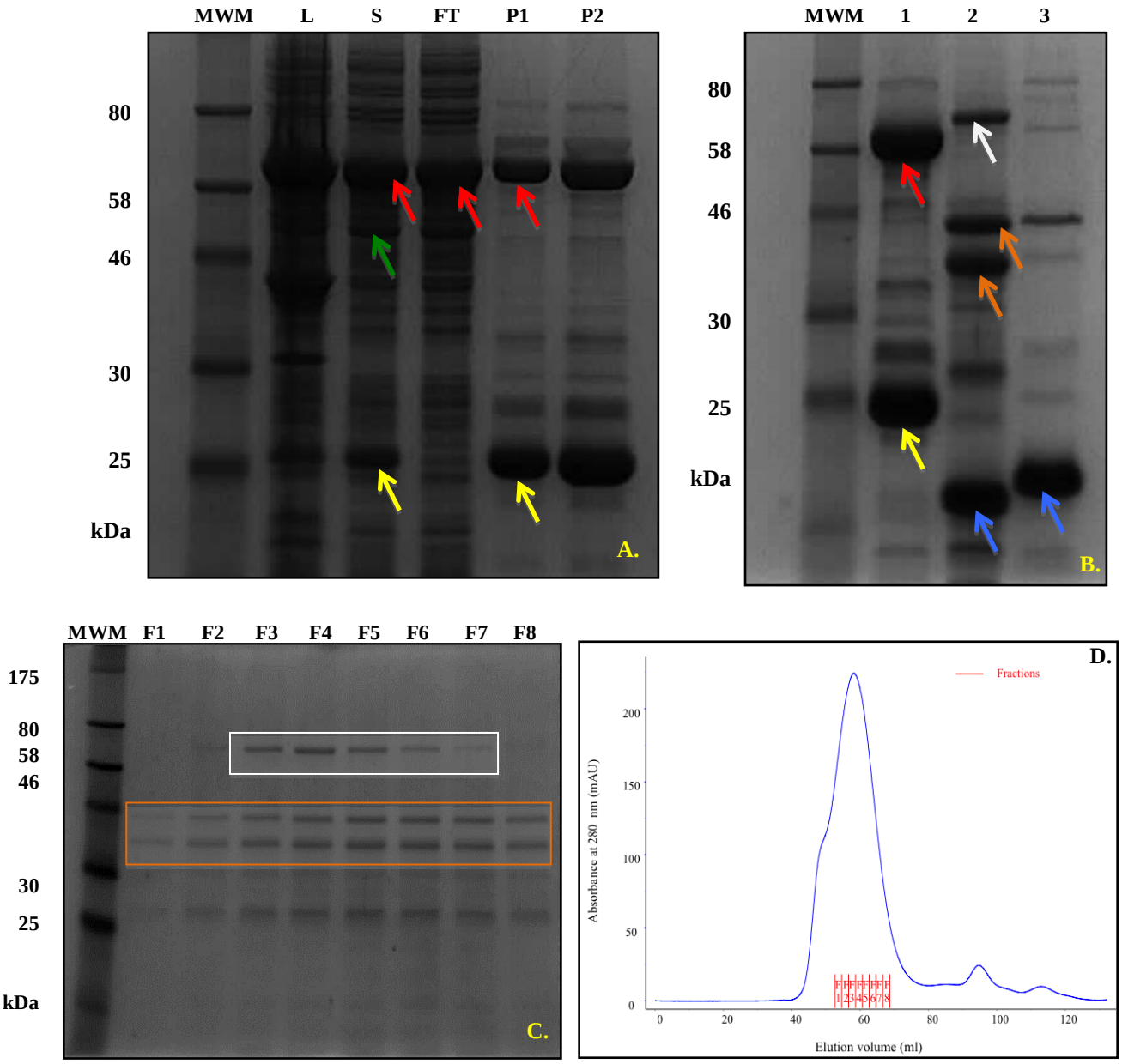


FIGURE AV.9 Expression and purification of Rv3038c orthologue in *M. thermoresistibile*

A. Metal affinity purification of SUMO tagged protein on Ni-sepharose column. Strong bands (red arrow) of full-length protein with SUMO tag seen at MW higher (MW of protein- 36.6 kDa, MW of SUMO tag-13.8 kDa, total MW of tagged protein-50.4 kDa). Due to column overloading, some of the tagged-protein flowed through. N-terminal part of protein with SUMO tag getting clipped seen at 25kDa (yellow arrows). Tag-less truncated protein left after clipping (green arrow) seen at 46 kDa. (MWM: molecular weight marker, L: cellular lysate, FT: flow through, S: supernatant, P1 to P2: protein).

B. Cleavage of SUMO tag by UlpI protease and reverse binding on Ni-sepharose column. 1: tagged protein, 2: tag-less protein obtained in the flow through, 3: SUMO tag. Due to column overloading, entire quantity of SUMO tag didn't get reversely bound and flowed through along with the tag-less truncated target protein. Full length protein with SUMO is marked by red arrow; N-terminal region with SUMO tag getting clipped during expression is marked by a yellow arrow; tag-less truncated protein left after UlpI cleavage is marked by orange arrows; SUMO tag is marked by blue arrow; chaperone GroEL is marked by white arrow.

C. Gel filtration of tag less protein on Superdex-200 column. Bands in white box: GroEL, bands in orange box: tag-less truncated protein after clipping during UlpI cleavage. (MWM: molecular weight marker, F1 to F8: protein fractions).

D. Gel filtration profile.

An alternate purification strategy was adopted to address the problem of clipping during expression. In this, affinity chromatography, treatment of full-length protein with EDTA and anionic exchange were performed. The clipping of a part of N-terminal conjoint to the SUMO tag during expression and thus, loss in yield of purified protein remained a problem, although the full-length tagged protein produced was pure enough.

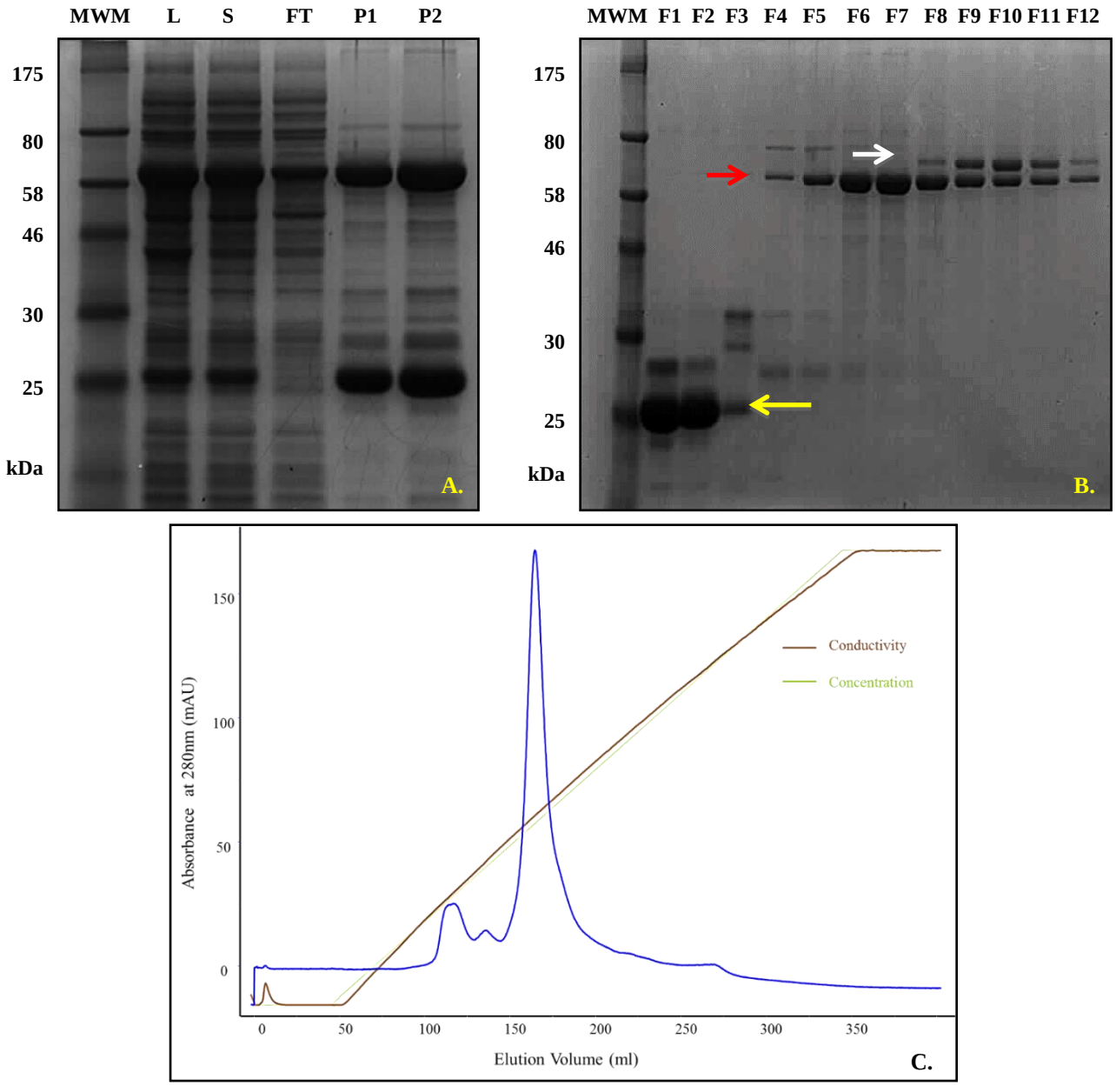


FIGURE AV.10 Expression and purification of SUMO-tagged Rv3038c orthologue in *M. thermoresistibile*

A. Metal affinity purification of SUMO tagged protein on Ni-sepharose column. (MWM-molecular weight marker, L-lysate, FT-flow through, S-supernatant, P1 to P2-protein)

B. Anionic exchange chromatography. Full length protein with SUMO tag (F4-F12) is marked by red arrow, N-terminal region with SUMO tag getting clipped (F1-F3) is marked by yellow arrow, chaperone GroEL (F8-F12, upper band) is marked by white arrow (MWM-molecular weight marker, F1 to F12-protein fractions)

C. Anionic exchange chromatography profile.

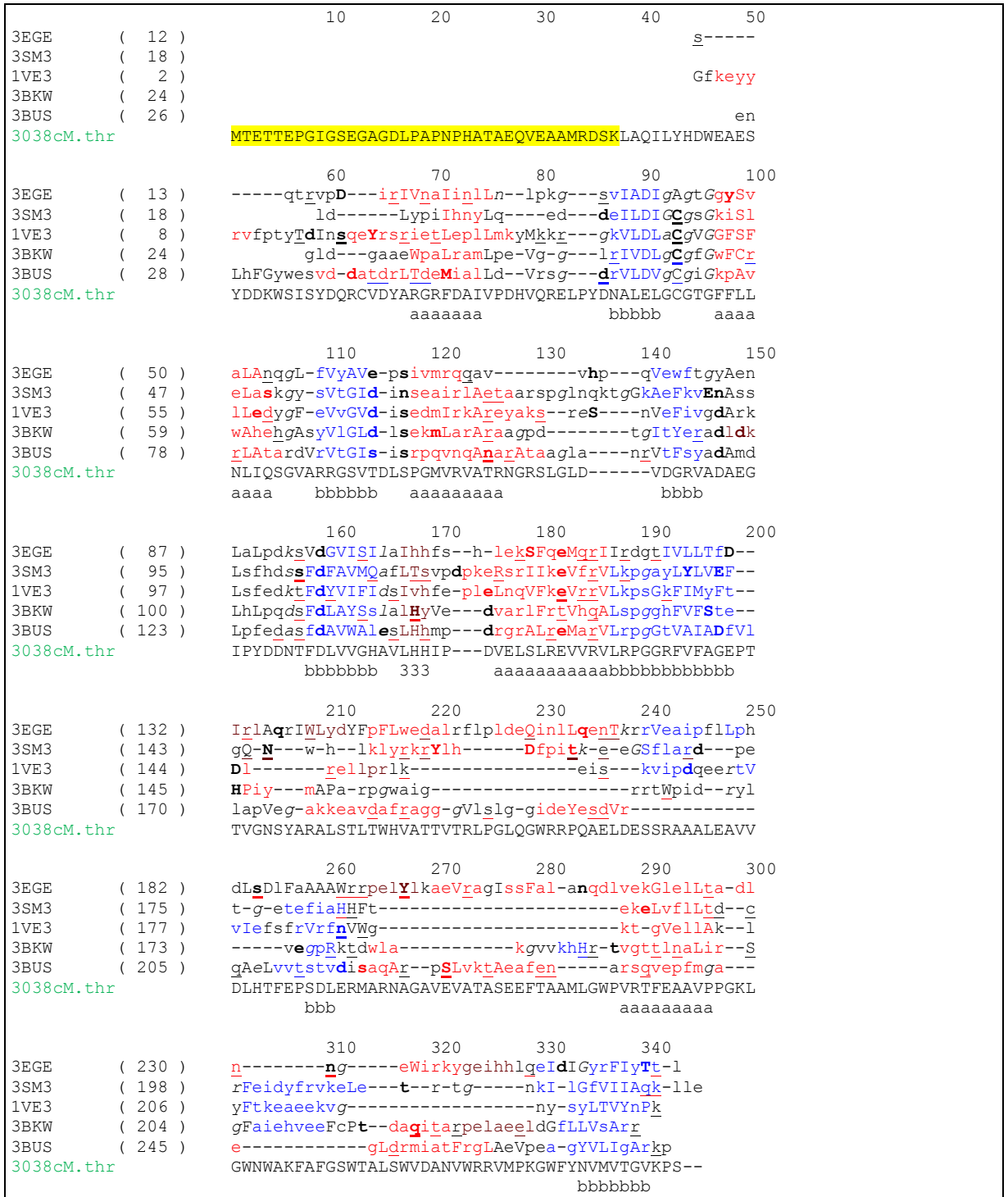
AV.3.2.4 Protein modeling of Rv3038c orthologue in *M. thermoresistibile*

The templates used for building model of orthologous protein of Rv3038c were methyltransferase proteins with structures available in PDB (PDB ID of templates- 3EGE, 3SM3, 1VE3, 3BKW, 3BUS).

PDB ID	PROTEIN	ORGANISM
1VE3_A	PH0226 protein	*Pyrococcus horikoshii OT3*
3BKW_B	S-adenosylmethionine dependent methyltransferase (NP_104914.1)	*Mesorhizobium loti MAFF303099*
3EGE_A	Putative methyltransferase from antibiotic biosynthesis pathway (YP_324569.1)	*Anabaena variabilis ATCC 29413*
3SM3_A	SAM-dependent methyltransferases Q8PUK2_METMA	*Methanosarcina mazei Go1*
3BUS_B	Rebeccamycin sugar 4'-O-methyltransferase RebM	*Lechevalieria aerocolonigenes*

Table AV.4 Structural templates used for homology modelling of orthologue of Rv3038c in *M. thermoresistibile.*

Model of the *M. thermoresistibile* protein indicated that the N-terminus region clipped is disordered. This is supported by the prediction with DisoPred. Based on these observations, this part of protein was truncated to retrieve the remaining protein.



Solvent inaccessible	UPPER CASE	X	hydrogen bond to main chain amide	**bold**	**x**
Solvent accessible	lowercase	x	hydrogen bond to main chain carbonyl	underline	x
α-helix	red	x	3_{10} helix	maroon	x
β-strand	blue	x	positive phi	*italic*	*x*

A.

B.

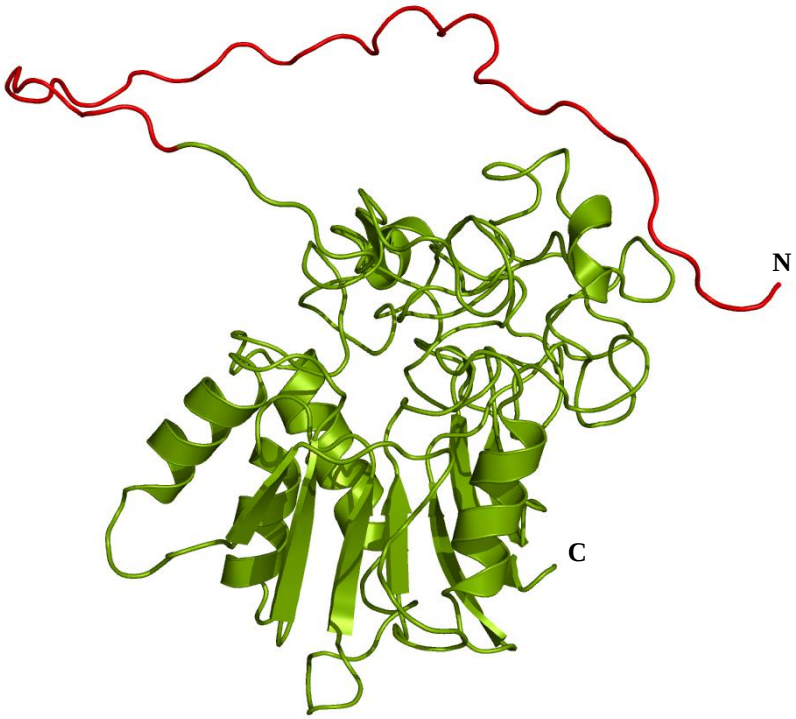


FIGURE AV.11 Model of three-dimensional structure of Rv3038c homologue in *M. thermoresistibile*

A. Structural alignment of the target protein and templates (produced by FUGUEALI). The part of N- terminus of the target protein that was clipped (highlighted in yellow) is not aligned to any secondary structure element. The methyltransferase domains of all the proteins align fairly well. (3038cM.thr refers to Rv3038c homologue in *M. thermoresistibile*)

B. Model of the target protein. Strand at the N-terminus of the protein getting naturally clipped (corresponding to the yellow highlighted part in the structural alignment above) is shaded red.

AV.3.2.5 N-terminal truncation of Rv3038c orthologue in *M. thermoresistibile*

Expression trial of SUMO-tagged truncated Rv3038c orthologue in *M. thermoresistibile*

The trial of truncated protein tagged with SUMO at its N-terminal illustrated that the protein was found in the pellet fraction and there was still clipping of a few residues from the N-terminus of the truncated protein. This was confirmed by MALDI fingerprinting of the bands.

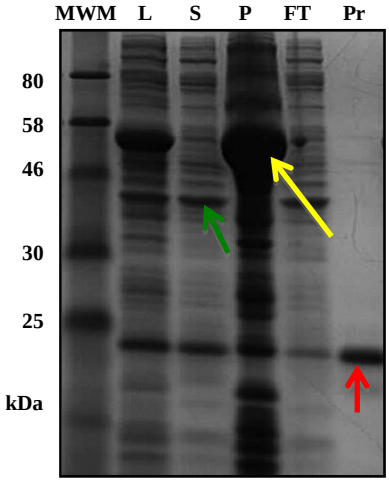


FIGURE AV.12 Expression trial of SUMO-tagged truncated Rv3038c orthologue in *M. thermoresistibile.*

Truncated SUMO tagged protein (yellow arrow) is seen in the pellet fraction. N-terminal part of the protein along with the SUMO tag is still getting clipped as seen at 25 kDa (red

Band observed at 46 kDa (green arrow) is of *E. coli* protein. (MWM: weight marker molecular, L: cellular lysate, S: supernatant, P: pellet, FT: flow through, Pr: protein)

Expression trial of His_6-tagged truncated Rv3038c orthologue in *M. thermoresistibile*

The trial of truncated protein tagged with His_6 at its C-terminal showed that the protein is found in inclusion bodies, although there was no clipping at N-terminal. A band observed in the soluble fraction was of a protein in *E. coli*, as shown by the MALDI fingerprinting of the bands.

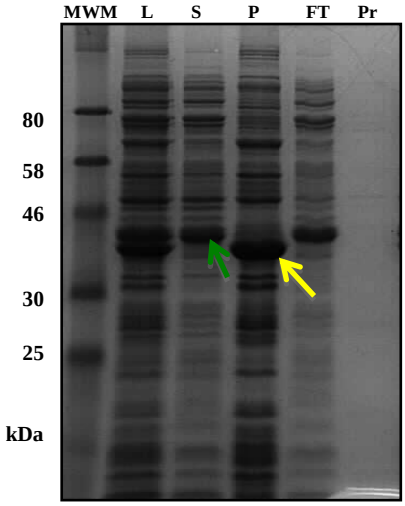


FIGURE AV.13 Expression trial of His_6-tagged truncated Rv3038c orthologue in *M. thermoresistibile*.

Truncated target protein (yellow arrow) is seen in the pellet fraction. Band observed in supernatant lane (green arrow) of *E. coli* protein.
(MWM: molecular weight marker, L: cellular lysate, S: supernatant, P: pellet, FT: flow through, Pr; protein)

BIBLIOGRAPHY

Aaron, L., Saadoun, D., Calatroni, I., Launay, O., Mémain, N., Vincent, V., Marchal, G., Dupont, B., Bouchaud, O., Valeyre, D., et al. (2004). Tuberculosis in HIV-infected patients: a comprehensive review. Clin. Microbiol. Infect. Off. Publ. Eur. Soc. Clin. Microbiol. Infect. Dis. *10*, 388–398.

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). *PHENIX* : a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. *66*, 213–221.

Alangaden, G.J., Kreiswirth, B.N., Aouad, A., Khetarpal, M., Igno, F.R., Moghazeh, S.L., Manavathu, E.K., and Lerner, S.A. (1998). Mechanism of resistance to amikacin and kanamycin in Mycobacterium tuberculosis. Antimicrob. Agents Chemother. *42*, 1295–1297.

Alangari, A.A., Al-Zamil, F., Al-Mazrou, A., Al-Muhsen, S., Boisson-Dupuis, S., Awadallah, S., Kambal, A., and Casanova, J.-L. (2011). Treatment of disseminated mycobacterial infection with high-dose IFN-γ in a patient with IL-12Rβ1 deficiency. Clin. Dev. Immunol. *2011*, 691956.

Alarico, S., Costa, M., Sousa, M.S., Maranha, A., Lourenço, E.C., Faria, T.Q., Ventura, M.R., and Empadinhas, N. (2014). Mycobacterium hassiacum recovers from nitrogen starvation with up-regulation of a novel glucosylglycerate hydrolase and depletion of the accumulated glucosylglycerate. Sci. Rep. *4*, 6766.

Alibaud, L., Pawelczyk, J., Gannoun-Zaki, L., Singh, V.K., Rombouts, Y., Drancourt, M., Dziadek, J., Guerardel, Y., and Kremer, L. (2014). Increased Phagocytosis of Mycobacterium marinum Mutants Defective in Lipooligosaccharide Production: A STRUCTURE-ACTIVITY RELATIONSHIP STUDY. J. Biol. Chem. *289*, 215–228.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Andries, K., Verhasselt, P., Guillemont, J., Göhlmann, H.W.H., Neefs, J.-M., Winkler, H., Van Gestel, J., Timmerman, P., Zhu, M., Lee, E., et al. (2005). A diarylquinoline drug active on the ATP synthase of Mycobacterium tuberculosis. Science *307*, 223–227.

Angala, S.K., Belardinelli, J.M., Huc-Claustre, E., Wheat, W.H., and Jackson, M. (2014). The cell envelope glycoconjugates of Mycobacterium tuberculosis. Crit. Rev. Biochem. Mol. Biol. *49*, 361–399.

Appelmelk, B.J., den Dunnen, J., Driessen, N.N., Ummels, R., Pak, M., Nigou, J., Larrouy-Maumus, G., Gurcha, S.S., Movahedzadeh, F., Geurtsen, J., et al. (2008). The mannose cap of mycobacterial lipoarabinomannan does not dominate the Mycobacterium-host interaction. Cell. Microbiol. *10*, 930–944.

Asención Diez, M.D., Demonte, A.M., Syson, K., Arias, D.G., Gorelik, A., Guerrero, S.A., Bornemann, S., and Iglesias, A.A. (2015). Allosteric regulation of the partitioning of glucose-1-phosphate between glycogen and trehalose biosynthesis in Mycobacterium tuberculosis. Biochim. Biophys. Acta BBA - Gen. Subj. *1850*, 13–21.

Ashkenazy, H., Erez, E., Martz, E., Pupko, T., and Ben-Tal, N. (2010). ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res. *38*, W529–W533.

Ashtekar, D.R., Costa-Periera, R., Shrinivasan, T., Iyyer, R., Vishvanathan, N., and Rittel, W. (1991). Oxazolidinones, a new class of synthetic antituberculosis agent. In vitro and in vivo activities of DuP-721 against Mycobacterium tuberculosis. Diagn. Microbiol. Infect. Dis. *14*, 465–471.

Avonce, N., Mendoza-Vargas, A., Morett, E., and Iturriaga, G. (2006). Insights on the evolution of trehalose biosynthesis. BMC Evol. Biol. *6*, 109.

Baba, T., Kaneda, K., Kusunose, E., Kusunose, M., and Yano, I. (1989). Thermally adaptive changes of mycolic acids in Mycobacterium smegmatis. J. Biochem. (Tokyo) *106*, 81–86.

Battye, T.G.G., Kontogiannis, L., Johnson, O., Powell, H.R., and Leslie, A.G.W. (2011). *iMOSFLM* : a new graphical interface for diffraction-image processing with *MOSFLM*. Acta Crystallogr. D Biol. Crystallogr. *67*, 271–281.

Baulard, A.R., Betts, J.C., Engohang-Ndong, J., Quan, S., McAdam, R.A., Brennan, P.J., Locht, C., and Besra, G.S. (2000). Activation of the pro-drug ethionamide is regulated in mycobacteria. J. Biol. Chem. *275*, 28326–28331.

Becker, A., Schlöder, P., Steele, J.E., and Wegener, G. (1996). The regulation of trehalose metabolism in insects. Experientia *52*, 433–439.

Belanger, A.E., Besra, G.S., Ford, M.E., Mikusová, K., Belisle, J.T., Brennan, P.J., and Inamine, J.M. (1996). The embAB genes of Mycobacterium avium encode an arabinosyl transferase involved in cell wall arabinan biosynthesis that is the target for the antimycobacterial drug ethambutol. Proc. Natl. Acad. Sci. U. S. A. *93*, 11919–11924.

Bell, W., Sun, W., Hohmann, S., Wera, S., Reinders, A., Virgilio, C.D., Wiemken, A., and Thevelein, J.M. (1998). Composition and Functional Analysis of the Saccharomyces cerevisiae Trehalose Synthase Complex. J. Biol. Chem. *273*, 33311–33319.

Benator, D., Bhattacharya, M., Bozeman, L., Burman, W., Cantazaro, A., Chaisson, R., Gordin, F., Horsburgh, C.R., Horton, J., Khan, A., et al. (2002). Rifapentine and isoniazid once a week versus rifampicin and isoniazid twice a week for treatment of drug-susceptible pulmonary tuberculosis in HIV-negative patients: a randomised clinical trial. Lancet Lond. Engl. *360*, 528–534.

Benvenuti, M., and Mangani, S. (2007). Crystallization of soluble proteins in vapor diffusion for x-ray crystallography. Nat. Protoc. *2*, 1633–1651.

Berg, S., Kaur, D., Jackson, M., and Brennan, P.J. (2007). The glycosyltransferases of Mycobacterium tuberculosis - roles in the synthesis of arabinogalactan, lipoarabinomannan, and other glycoconjugates. Glycobiology *17*, 35–56R.

Bergfors, T.M. (2009). Protein Crystallization, 2nd Ed. (Internat'l University Line).

Besra, G.S., and Brennan, P.J. (1997). The mycobacterial cell wall: biosynthesis of arabinogalactan and lipoarabinomannan. Biochem. Soc. Trans. *25*, 845–850.

Besra, G.S., McNeil, M.R., and Brennan, P.J. (1992). Characterization of the specific antigenicity of Mycobacterium fortuitum. Biochemistry (Mosc.) *31*, 6504–6509.

Bhamidi, S., Scherman, M.S., Rithner, C.D., Prenni, J.E., Chatterjee, D., Khoo, K.-H., and McNeil, M.R. (2008). The Identification and Location of Succinyl Residues and the Characterization of the Interior Arabinan Region Allow for a Model of the Complete Primary Structure of Mycobacterium tuberculosis Mycolyl Arabinogalactan. J. Biol. Chem. *283*, 12992–13000.

Bloch, K. (1977). Control mechanisms for fatty acid synthesis in Mycobacterium smegmatis. Adv. Enzymol. Relat. Areas Mol. Biol. *45*, 1–84.

Blundell, T.L., and Johnson, L.N. (1976). Protein Crystallography (Academic Press).

Bollag, G., Hirth, P., Tsai, J., Zhang, J., Ibrahim, P.N., Cho, H., Spevak, W., Zhang, C., Zhang, Y., Habets, G., et al. (2010). Clinical efficacy of a RAF inhibitor needs broad target blockade in BRAF-mutant melanoma. Nature *467*, 596–599.

Bollela, V.R., Namburete, E.I., Feliciano, C.S., Macheque, D., Harrison, L.H., and Caminero, J.A. (2016). Detection of katG and inhA mutations to guide isoniazid and ethionamide use for drug-resistant tuberculosis. Int. J. Tuberc. Lung Dis. Off. J. Int. Union Tuberc. Lung Dis. *20*, 1099–1104.

Bourne, Y., and Henrissat, B. (2001). Glycoside hydrolases and glycosyltransferases: families and functional modules. Curr. Opin. Struct. Biol. *11*, 593–600.

Breen, R.A.M., Swaden, L., Ballinger, J., and Lipman, M.C.I. (2006). Tuberculosis and HIV co-infection: a practical therapeutic approach. Drugs *66*, 2299–2308.

Brennan, P.. (2003). Structure, function, and biogenesis of the cell wall of Mycobacterium tuberculosis. Tuberculosis *83*, 91–97.

Briken, V., Porcelli, S.A., Besra, G.S., and Kremer, L. (2004). Mycobacterial lipoarabinomannan and related lipoglycans: from biogenesis to modulation of the immune response. Mol. Microbiol. *53*, 391–403.

Bruning, J.B., Murillo, A.C., Chacon, O., Barletta, R.G., and Sacchettini, J.C. (2011). Structure of the Mycobacterium tuberculosis d-Alanine:d-Alanine Ligase, a Target of the Antituberculosis Drug d-Cycloserine. Antimicrob. Agents Chemother. *55*, 291–301.

Bujnicki, J.M. (2000). Comparison of Protein Structures Reveals Monophyletic Origin of the AdoMet-Dependent Methyltransferase Family and Mechanistic Convergence Rather than Recent Differentiation of N4-Cytosine and N6-Adenine DNA Methylation. In Silico Biol. *1*, 175–182.

Cabib, E., and Leloir, L.F. (1958). The biosynthesis of trehalose phosphate. J. Biol. Chem. *231*, 259–275.

Campbell, J.A., Davies, G.J., Bulone, V., and Henrissat, B. (1997). A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. Biochem. J. *326*, 929.

Candy, D.J., and Kilby, B.A. (1959). Site and mode of trehalose biosynthesis in the locust. Nature *183*, 1594–1595.

Caner, S., Nguyen, N., Aguda, A., Zhang, R., Pan, Y.T., Withers, S.G., and Brayer, G.D. (2013). The structure of the Mycobacterium smegmatis trehalose synthase reveals an unusual active site configuration and acarbose-binding mode. Glycobiology *23*, 1075–1083.

Cantaloube, S., Veyron-Churlet, R., Haddache, N., Daffé, M., and Zerbib, D. (2011). The Mycobacterium Tuberculosis FAS-II Dehydratases and Methyltransferases Define the Specificity of the Mycolic Acid Elongation Complexes. PLoS ONE *6*, e29564.

Carroll, J.D., Pastuszak, I., Edavana, V.K., Pan, Y.T., and Elbein, A.D. (2007). A novel trehalase from Mycobacterium smegmatis - purification, properties, requirements. FEBS J. *274*, 1701–1714.

Cason, J., Allen, C.F., Deacetis, W., and Fonken, G.J. (1956). Fatty acids from the lipides of non-virulent strains of the tubercle bacillus. J. Biol. Chem. *220*, 893–904.

Cavalier, M.C., Yim, Y.-S., Asamizu, S., Neau, D., Almabruk, K.H., Mahmud, T., and Lee, Y.-H. (2012). Mechanistic Insights into Validoxylamine A 7'-Phosphate Synthesis by VldE Using the Structure of the Entire Product Complex. PLoS ONE *7*, e44934.

Centers for Disease Control and Prevention (CDC) (2002). Acquired rifamycin resistance in persons with advanced HIV disease being treated for active tuberculosis with intermittent rifamycin-based regimens. MMWR Morb. Mortal. Wkly. Rep. *51*, 214–215.

Chan, E.D., and Iseman, M.D. (2008). Multidrug-resistant and extensively drug-resistant tuberculosis: a review. Curr. Opin. Infect. Dis. *21*, 587–595.

Chatterjee, D., and Khoo, K.H. (1998). Mycobacterial lipoarabinomannan: an extraordinary lipoheteroglycan with profound physiological effects. Glycobiology *8*, 113–120.

Chatterjee, D., Hunter, S.W., McNeil, M., and Brennan, P.J. (1992). Lipoarabinomannan. Multiglycosylated form of the mycobacterial mannosylphosphatidylinositols. J. Biol. Chem. *267*, 6228–6233.

Chen, Q., Ma, E., Behar, K.L., Xu, T., and Haddad, G.G. (2002). Role of trehalose phosphate synthase in anoxia tolerance and development in Drosophila melanogaster. J. Biol. Chem. *277*, 3274–3279.

Cheng, X., and Roberts, R.J. (2001). AdoMet-dependent methylation, DNA methyltransferases and base flipping. Nucleic Acids Res. *29*, 3784–3795.

Cheung, M.-S., Maguire, M.L., Stevens, T.J., and Broadhurst, R.W. (2010). DANGLE: A Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure. J. Magn. Reson. *202*, 223–233.

Chien, J.-Y., Chiu, W.-Y., Chien, S.-T., Chiang, C.-J., Yu, C.-J., and Hsueh, P.-R. (2016). Mutations in gyrA and gyrB among Fluoroquinolone- and Multidrug-Resistant Mycobacterium tuberculosis Isolates. Antimicrob. Agents Chemother. *60*, 2090–2096.

Cole, St., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry, C. 3rd, et al. (1998). Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature *393*, 537–544.

Cole, S.T., Eiglmeier, K., Parkhill, J., James, K.D., Thomson, N.R., Wheeler, P.R., Honoré, N., Garnier, T., Churcher, C., Harris, D., et al. (2001). Massive gene decay in the leprosy bacillus. Nature *409*, 1007–1011.

Cook, J.L. (2010). Nontuberculous mycobacteria: opportunistic environmental pathogens for predisposed hosts. Br. Med. Bull. *96*, 45–59.

Coutinho, P.M., Deleury, E., Davies, G.J., and Henrissat, B. (2003). An Evolving Hierarchical Family Classification for Glycosyltransferases. J. Mol. Biol. *328*, 307–317.

Cox, E., and Laessig, K. (2014). FDA approval of bedaquiline--the benefit-risk balance for drug-resistant tuberculosis. N. Engl. J. Med. *371*, 689–691.

Cywes, C., Hoppe, H.C., Daffé, M., and Ehlers, M.R. (1997). Nonopsonic binding of Mycobacterium tuberculosis to complement receptor type 3 is mediated by capsular polysaccharides and is strain dependent. Infect. Immun. *65*, 4258–4266.

Daffé, M., Lacave, C., Lanéelle, M.A., Gillois, M., and Lanéelle, G. (1988). Polyphthienoyl trehalose, glycolipids specific for virulent strains of the tubercle bacillus. Eur. J. Biochem. FEBS *172*, 579–584.

Daffe, M., McNeil, M., and Brennan, P.J. (1991). Novel type-specific lipooligosaccharides from Mycobacterium tuberculosis. Biochemistry (Mosc.) *30*, 378–388.

Daniel, T.M. (2006). The history of tuberculosis. Respir. Med. *100*, 1862–1870.

Davies, T.G., and Hyvönen, M. (2012). Fragment-Based Drug Discovery and X-Ray Crystallography (Springer Science & Business Media).

De Lorenzo, S., Alffenaar, J.W., Sotgiu, G., Centis, R., D'Ambrosio, L., Tiberi, S., Bolhuis, M.S., van Altena, R., Viggiani, P., Piana, A., et al. (2013). Efficacy and safety of meropenem-clavulanate added to linezolid-containing regimens in the treatment of MDR-/XDR-TB. Eur. Respir. J. *41*, 1386–1392.

De Smet, K.A., Weston, A., Brown, I.N., Young, D.B., and Robertson, B.D. (2000). Three pathways for trehalose biosynthesis in mycobacteria. Microbiology *146*, 199–208.

DeBarber, A.E., Mdluli, K., Bosman, M., Bekker, L.G., and Barry, C.E. (2000). Ethionamide activation and sensitivity in multidrug-resistant Mycobacterium tuberculosis. Proc. Natl. Acad. Sci. U. S. A. *97*, 9677–9682.

Deller, M.C., Kong, L., and Rupp, B. (2016). Protein stability: a crystallographer's perspective. Acta Crystallogr. Sect. F Struct. Biol. Commun. *72*, 72–95.

Dey, T., Brigden, G., Cox, H., Shubber, Z., Cooke, G., and Ford, N. (2013). Outcomes of clofazimine for the treatment of drug-resistant tuberculosis: a systematic review and meta-analysis. J. Antimicrob. Chemother. *68*, 284–293.

Dhariwal, K.R., Yang, Y.M., Fales, H.M., and Goren, M.B. (1987). Detection of trehalose monomycolate in Mycobacterium leprae grown in armadillo tissues. J. Gen. Microbiol. *133*, 201–209.

Diacon, A.H., Pym, A., Grobusch, M., Patientia, R., Rustomjee, R., Page-Shipp, L., Pistorius, C., Krause, R., Bogoshi, M., Churchyard, G., et al. (2009). The diarylquinoline TMC207 for multidrug-resistant tuberculosis. N. Engl. J. Med. *360*, 2397–2405.

Diacon, A.H., Dawson, R., Hanekom, M., Narunsky, K., Venter, A., Hittel, N., Geiter, L.J., Wells, C.D., Paccaly, A.J., and Donald, P.R. (2011). Early bactericidal activity of delamanid (OPC-67683) in smear-positive pulmonary tuberculosis patients. Int. J. Tuberc. Lung Dis. Off. J. Int. Union Tuberc. Lung Dis. *15*, 949–954.

Dickinson, J.M., and Mitchison, D.A. (1981). Experimental models to explain the high sterilizing activity of rifampin in the chemotherapy of tuberculosis. Am. Rev. Respir. Dis. *123*, 367–371.

Doak, B.C., Norton, R.S., and Scanlon, M.J. (2016). The ways and means of fragment-based drug design. Pharmacol. Ther.

Donald, P.R., and Diacon, A.H. (2015). Para-aminosalicylic acid: the return of an old friend. Lancet Infect. Dis. *15*, 1091–1099.

Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.-H., and Stackebrandt, E. (2006). The Prokaryotes (New York, NY: Springer New York).

Dye, C. (2009). Doomsday postponed? Preventing and reversing epidemics of drug-resistant tuberculosis. Nat. Rev. Microbiol. *7*, 81–87.

Edavana, V.K., Pastuszak, I., Carroll, J.., Thampi, P., Abraham, E.C., and Elbein, A.D. (2004). Cloning and expression of the trehalose-phosphate phosphatase of Mycobacterium tuberculosis: comparison to the enzyme from Mycobacterium smegmatis. Arch. Biochem. Biophys. *426*, 250–257.

Edwards, T.E., Liao, R., Phan, I., Myler, P.J., and Grundner, C. (2012). Mycobacterium thermoresistibile as a source of thermostable orthologs of Mycobacterium tuberculosis proteins. Protein Sci. *21*, 1093–1096.

Ehlers, M.R., and Daffé, M. (1998). Interactions between Mycobacterium tuberculosis and host cells: are mycobacterial sugars the key? Trends Microbiol. *6*, 328–335.

Elbein, A.D. (2003). New insights on trehalose: a multifunctional molecule. Glycobiology *13*, 17R–27.

Elbein, A.D., Pastuszak, I., Tackett, A.J., Wilson, T., and Pan, Y.T. (2010). Last step in the conversion of trehalose to glycogen: a mycobacterial enzyme that transfers maltose from maltose 1-phosphate to glycogen. J. Biol. Chem. *285*, 9803–9812.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of *Coot*. Acta Crystallogr. D Biol. Crystallogr. *66*, 486–501.

Errey, J.C., Lee, S.S., Gibson, R.P., Martinez Fleites, C., Barry, C.S., Jung, P.M.J., O'Sullivan, A.C., Davis, B.G., and Davies, G.J. (2010). Mechanistic Insight into Enzymatic Glycosyl Transfer with Retention of Configuration through Analysis of Glycomimetic Inhibitors. Angew. Chem. Int. Ed. *49*, 1234–1237.

Etienne, G., Malaga, W., Laval, F., Lemassu, A., Guilhot, C., and Daffe, M. (2009). Identification of the Polyketide Synthase Involved in the Biosynthesis of the Surface-Exposed Lipooligosaccharides in Mycobacteria. J. Bacteriol. *191*, 2613–2621.

Evans, P.R., and Murshudov, G.N. (2013). How good are my data and what is the resolution? Acta Crystallogr. D Biol. Crystallogr. *69*, 1204–1214.

Falkinham, J.O. (2009). Surrounded by mycobacteria: nontuberculous mycobacteria in the human environment. J. Appl. Microbiol. *107*, 356–367.

Ferguson, J.A., and Ballou, C.E. (1970). Biosynthesis of a Mycobacterial Lipopolysaccharide PROPERTIES OF THE POLYSACCHARIDE METHYLTRANSFERASE. J. Biol. Chem. *245*, 4213–4223.

Finken, M., Kirschner, P., Meier, A., Wrede, A., and Böttger, E.C. (1993). Molecular basis of streptomycin resistance in Mycobacterium tuberculosis: alterations of the ribosomal protein S12 gene and point mutations within a functional 16S ribosomal RNA pseudoknot. Mol. Microbiol. *9*, 1239–1246.

Flick, P.K., and Bloch, K. (1974). In vitro alterations of the product distribution of the fatty acid synthetase from Mycobacterium phlei. J. Biol. Chem. *249*, 1031–1036.

Fortún, J., Martín-Dávila, P., Navas, E., Pérez-Elías, M.J., Cobo, J., Tato, M., De la Pedrosa, E.G.-G., Gómez-Mampaso, E., and Moreno, S. (2005). Linezolid for the treatment of multidrug-resistant tuberculosis. J. Antimicrob. Chemother. *56*, 180–185.

Freyer, M.W., and Lewis, E.A. (2008). Isothermal Titration Calorimetry: Experimental Design, Data Analysis, and Probing Macromolecule/Ligand Binding and Kinetic Interactions. In Methods in Cell Biology, (Elsevier), pp. 79–113.

Frith, J. (2014). History of tuberculosis. Part 1-phthisis, consumption and the white plague.

Gancedo, C., and Flores, C. (2004). The importance of a functional trehalose biosynthetic pathway for the life of yeasts and fungi. FEMS Yeast Res. *4*, 351–359.

Gandhi, N.R., Nunn, P., Dheda, K., Schaaf, H.S., Zignol, M., van Soolingen, D., Jensen, P., and Bayona, J. (2010). Multidrug-resistant and extensively drug-resistant tuberculosis: a threat to global control of tuberculosis. Lancet Lond. Engl. *375*, 1830–1843.

Gangadharam, P.R., Cohn, M.L., and Middlebrook, G. (1963). INFECTIVITY, PATHOGENICITY AND SULPHOLIPID FRACTION OF SOME INDIAN AND BRITISH STRAINS OF TUBERCLE BACILLI. Tubercle *44*, 452–455.

Garg, S.K., Alam, M.S., Kishan, K.V.R., and Agrawal, P. (2007). Expression and characterization of alpha-(1,4)-glucan branching enzyme Rv1326c of Mycobacterium tuberculosis H37Rv. Protein Expr. Purif. *51*, 198–208.

Garman, E.F. (2014). Developments in X-ray Crystallographic Structure Determination of Biological Macromolecules. Science *343*, 1102–1108.

Gasteiger, E. (2003). ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. *31*, 3784–3788.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S. 'everine, Wilkins, M.R., Appel, R.D., and Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server (Springer).

Gautier, N., López Marín, L.M., Lanéelle, M.A., and Daffé, M. (1992). Structure of mycoside F, a family of trehalose-containing glycolipids of Mycobacterium fortuitum. FEMS Microbiol. Lett. *77*, 81–87.

Gibson, R.P., Turkenburg, J.P., Charnock, S.J., Lloyd, R., and Davies, G.J. (2002a). Insights into trehalose synthesis provided by the structure of the retaining glucosyltransferase OtsA. Chem. Biol. *9*, 1337–1346.

Gibson, R.P., Lloyd, R.M., Charnock, S.J., and Davies, G.J. (2002b). Characterization of Escherichia coli OtsA, a trehalose-6-phosphate synthase from glycosyltransferase family 20. Acta Crystallogr. D Biol. Crystallogr. *58*, 349–351.

Gibson, R.P., Tarling, C.A., Roberts, S., Withers, S.G., and Davies, G.J. (2004). The Donor Subsite of Trehalose-6-phosphate Synthase: BINARY COMPLEXES WITH UDP-GLUCOSE AND UDP-2-DEOXY-2-FLUORO-GLUCOSE AT 2 A RESOLUTION. J. Biol. Chem. *279*, 1950–1955.

Giegé, R. (2013). A historical perspective on protein crystallization from 1840 to the present day. FEBS J. *280*, 6456–6497.

Gilleron, M., Bala, L., Brando, T., Vercellone, A., and Puzo, G. (2000). Mycobacterium tuberculosis H37Rv parietal and cellular lipoarabinomannans. Characterization of the acyl- and glyco-forms. J. Biol. Chem. *275*, 677–684.

Gilleron, M., Ronet, C., Mempel, M., Monsarrat, B., Gachelin, G., and Puzo, G. (2001). Acylation state of the phosphatidylinositol mannosides from Mycobacterium bovis bacillus Calmette Guérin and ability to induce granuloma and recruit natural killer T cells. J. Biol. Chem. *276*, 34896–34904.

Gilleron, M., Quesniaux, V.F.J., and Puzo, G. (2003). Acylation state of the phosphatidylinositol hexamannosides from Mycobacterium bovis bacillus Calmette Guerin and mycobacterium tuberculosis H37Rv and its implication in Toll-like receptor response. J. Biol. Chem. *278*, 29880–29889.

Gilmore, S.A., Schelle, M.W., Holsclaw, C.M., Leigh, C.D., Jain, M., Cox, J.S., Leary, J.A., and Bertozzi, C.R. (2012). Sulfolipid-1 Biosynthesis Restricts *Mycobacterium tuberculosis* Growth in Human Macrophages. ACS Chem. Biol. *7*, 863–870.

Gler, M.T., Skripconoka, V., Sanchez-Garavito, E., Xiao, H., Cabrera-Rivero, J.L., Vargas-Vasquez, D.E., Gao, M., Awad, M., Park, S.-K., Shim, T.S., et al. (2012). Delamanid for multidrug-resistant pulmonary tuberculosis. N. Engl. J. Med. *366*, 2151–2160.

Global tuberculosis report, WHO (2015). Global tuberculosis report 2015 (Geneva: World Health Organization).

Goren, M.B. (1972). Mycobacterial lipids: selected topics. Bacteriol. Rev. *36*, 33–64.

Goren, M.B., D'Arcy Hart, P., Young, M.R., and Armstrong, J.A. (1976). Prevention of phagosome-lysosome fusion in cultured macrophages by sulfatides of Mycobacterium tuberculosis. Proc. Natl. Acad. Sci. U. S. A. *73*, 2510–2514.

Goude, R., Amin, A.G., Chatterjee, D., and Parish, T. (2009). The arabinosyltransferase EmbC is inhibited by ethambutol in Mycobacterium tuberculosis. Antimicrob. Agents Chemother. *53*, 4138–4146.

Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J., and Lopez, R. (2010). A new bioinformatics analysis tools framework at EMBL-EBI. Nucleic Acids Res. *38*, W695–W699.

Graewert, M.A., and Svergun, D.I. (2013). Impact and progress in small and wide angle X-ray scattering (SAXS and WAXS). Curr. Opin. Struct. Biol. *23*, 748–754.

Gray, G.R., and Ballou, C.E. (1971). Isolation and characterization of a polysaccharide containing 3-O-methyl-D-mannose from Mycobacterium phlei. J. Biol. Chem. *246*, 6835–6842.

Grellert, E., and Ballou, C.E. (1972). Biosynthesis of a Mycobacterial Lipopolysaccharide EVIDENCE FOR AN ACYLPOLYSACCHARIDE METHYLTRANSFERASE. J. Biol. Chem. *247*, 3236–3241.

Grennan, A.K. (2007). The Role of Trehalose Biosynthesis in Plants. PLANT Physiol. *144*, 3–5.

Griffin, J.E., Gawronski, J.D., Dejesus, M.A., Ioerger, T.R., Akerley, B.J., and Sassetti, C.M. (2011). High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. PLoS Pathog. *7*, e1002251.

Grover, S., Gupta, P., Kahlon, P.S., Goyal, S., Grover, A., Dalal, K., Sabeeha, S., Ehtesham, N.Z., and Hasnain, S.E. (2016). Analyses of methyltransferases across the pathogenicity spectrum of different mycobacterial species point to an extremophile connection. Mol BioSyst.

Harland, C.W., Rabuka, D., Bertozzi, C.R., and Parthasarathy, R. (2008). The Mycobacterium tuberculosis virulence factor trehalose dimycolate imparts desiccation resistance to model mycobacterial membranes. Biophys. J. *94*, 4718–4724.

Harris, L.S., and Gray, G.R. (1977). Acetylated methylmannose polysaccharide of Streptomyces. J. Biol. Chem. *252*, 2470–2477.

Hass, F., and Hass, S.S. (1996). The origins of Mycobacterium tuberculosis and the notion of its contagiousness. Rom WN Garay SM Eds Tuberc.

Heinrich, N., Dawson, R., du Bois, J., Narunsky, K., Horwith, G., Phipps, A.J., Nacy, C.A., Aarnoutse, R.E., Boeree, M.J., Gillespie, S.H., et al. (2015). Early phase evaluation of SQ109 alone and in combination with rifampicin in pulmonary TB patients. J. Antimicrob. Chemother. *70*, 1558–1566.

Hett, E.C., and Rubin, E.J. (2008). Bacterial Growth and Cell Division: a Mycobacterial Perspective. Microbiol. Mol. Biol. Rev. *72*, 126–156.

Hingley-Wilson, S.M., Sambandamurthy, V.K., and Jacobs, W.R. (2003). Survival perspectives from the world's most successful pathogen, Mycobacterium tuberculosis. Nat. Immunol. *4*, 949–955.

Holm, L., and Rosenstrom, P. (2010). Dali server: conservation mapping in 3D. Nucleic Acids Res. *38*, W545–W549.

Horowitz, S., Dirk, L.M.A., Yesselman, J.D., Nimtz, J.S., Adhikari, U., Mehl, R.A., Scheiner, S., Houtz, R.L., Al-Hashimi, H.M., and Trievel, R.C. (2013). Conservation and Functional Importance of Carbon–Oxygen Hydrogen Bonding in AdoMet-Dependent Methyltransferases. J. Am. Chem. Soc. *135*, 15536–15548.

Hottiger, T., Boller, T., and Wiemken, A. (1987). Rapid changes of heat and desiccation tolerance correlated with changes of trehalose content in Saccharomyces cerevisiae cells subjected to temperature shifts. FEBS Lett. *220*, 113–115.

Hugonnet, J.-E., Tremblay, L.W., Boshoff, H.I., Barry, C.E., and Blanchard, J.S. (2009). Meropenem-clavulanate is effective against extensively drug-resistant Mycobacterium tuberculosis. Science *323*, 1215–1218.

Huitric, E., Verhasselt, P., Koul, A., Andries, K., Hoffner, S., and Andersson, D.I. (2010). Rates and mechanisms of resistance development in Mycobacterium tuberculosis to a novel diarylquinoline ATP synthase inhibitor. Antimicrob. Agents Chemother. *54*, 1022–1028.

Hunter, S.W., Murphy, R.C., Clay, K., Goren, M.B., and Brennan, P.J. (1983). Trehalose-containing lipooligosaccharides. A new class of species-specific antigens from Mycobacterium. J. Biol. Chem. *258*, 10481–10487.

Hunter, S.W., Barr, V.L., McNeil, M., Jardine, I., and Brennan, P.J. (1988). Trehalose-containing lipooligosaccharide antigens of Mycobacterium sp.: presence of a mono-O-methyltri-O-acyltrehalose "core." Biochemistry (Mosc.) *27*, 1549–1556.

Ilton, M., Jevans, A.W., McCarthy, E.D., Vance, D., White, H.B., and Bloch, K. (1971). Fatty acid synthetase activity in Mycobacterium phlei: regulation by polysaccharides. Proc. Natl. Acad. Sci. U. S. A. *68*, 87–91.

Jackson, M., and Brennan, P.J. (2009). Polymethylated Polysaccharides from Mycobacterium Species Revisited. J. Biol. Chem. *284*, 1949–1953.

Jackson, M., Stadthagen, G., and Gicquel, B. (2007). Long-chain multiple methyl-branched fatty acid-containing lipids of Mycobacterium tuberculosis: biosynthesis, transport, regulation and biological activities. Tuberc. Edinb. Scotl. *87*, 78–86.

Jones, D.T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol. *292*, 195–202.

Kalscheuer, R., Syson, K., Veeraraghavan, U., Weinrick, B., Biermann, K.E., Liu, Z., Sacchettini, J.C., Besra, G., Bornemann, S., and Jacobs, W.R. (2010). Self-poisoning of Mycobacterium tuberculosis by targeting GlgE in an α-glucan pathway. Nat. Chem. Biol. *6*, 376–384.

Kamisango, K., Dell, A., and Ballou, C.E. (1987). Biosynthesis of the mycobacterial O-methylglucose lipopolysaccharide. Characterization of putative intermediates in the initiation, elongation, and termination reactions. J. Biol. Chem. *262*, 4580–4586.

Kapopoulou, A., Lew, J.M., and Cole, S.T. (2011). The MycoBrowser portal: A comprehensive and manually annotated resource for mycobacterial genomes. Tuberculosis *91*, 8–13.

Kato, M., and Goren, M.B. (1974). Synergistic action of cord factor and mycobacterial sulfatides on mitochondria. Infect. Immun. *10*, 733–741.

Kaufmann, S.H. (2001). How can immunology contribute to the control of tuberculosis? Nat. Rev. Immunol. *1*, 20–30.

Kaur, D., Guerin, M.E., Škovierová, H., Brennan, P.J., and Jackson, M. (2009a). Chapter 2 Biogenesis of the Cell Wall and Other Glycoconjugates of Mycobacterium tuberculosis. In Advances in Applied Microbiology, (Elsevier), pp. 23–78.

Kaur, D., Pham, H., Larrouy-Maumus, G., Rivière, M., Vissa, V., Guerin, M.E., Puzo, G., Brennan, P.J., and Jackson, M. (2009b). Initiation of Methylglucose Lipopolysaccharide Biosynthesis in Mycobacteria. PLoS ONE *4*, e5447.

Kawaguchi, A., and Bloch, K. (1974). Inhibition of glucose 6-phosphate dehydrogenase by palmitoyl coenzyme A. J. Biol. Chem. *249*, 5793–5800.

Keller, J.M., and Ballou, C.E. (1968). The 6-O-Methylglucose-containing Lipopolysaccharide of Mycobacterium phlei IDENTIFICATION OF THE LIPID COMPONENTS. J. Biol. Chem. *243*, 2905–2910.

Kern, C., Wolf, C., Bender, F., Berger, M., Noack, S., Schmalz, S., and Ilg, T. (2012). Trehalose-6-phosphate synthase from the cat flea Ctenocephalides felis and Drosophila melanogaster: gene identification, cloning, heterologous functional expression and identification of inhibitors by high throughput screening: Insect trehalose-6-phosphate synthase. Insect Mol. Biol. *21*, 456–471.

Kikhney, A.G., and Svergun, D.I. (2015). A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. FEBS Lett. *589*, 2570–2577.

Killick, K.A. (1979). Trehalose 6-phosphate synthase from Dictyostelium discoideum: partial purification and characterization of the enzyme from young sorocarps. Arch. Biochem. Biophys. *196*, 121–133.

Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J., and Svergun, D.I. (2003). *PRIMUS* : a Windows PC-based system for small-angle scattering data analysis. J. Appl. Crystallogr. *36*, 1277–1282.

Koul, A., Dendouga, N., Vergauwen, K., Molenberghs, B., Vranckx, L., Willebrords, R., Ristic, Z., Lill, H., Dorange, I., Guillemont, J., et al. (2007). Diarylquinolines target subunit c of mycobacterial ATP synthase. Nat. Chem. Biol. *3*, 323–324.

Koul, A., Vranckx, L., Dendouga, N., Balemans, W., Van den Wyngaert, I., Vergauwen, K., Göhlmann, H.W.H., Willebrords, R., Poncelet, A., Guillemont, J., et al. (2008). Diarylquinolines are bactericidal for dormant mycobacteria as a result of disturbed ATP homeostasis. J. Biol. Chem. *283*, 25273–25280.

Koul, A., Arnoult, E., Lounis, N., Guillemont, J., and Andries, K. (2011). The challenge of new drug discovery for tuberculosis. Nature *469*, 483–490.

Kovermann, M., Rogne, P., and Wolf-Watz, M. (2016). Protein dynamics and function from solution state NMR spectroscopy. Q. Rev. Biophys. *49*.

Kozbial, P.Z., and Mushegian, A.R. (2005). Natural history of S-adenosylmethionine-binding proteins. BMC Struct. Biol. *5*, 1.

Kranz, J.K., and Schalk-Hihi, C. (2011). Protein Thermal Shifts to Identify Low Molecular Weight Fragments. In Methods in Enzymology, (Elsevier), pp. 277–298.

Kumar, P., Arora, K., Lloyd, J.R., Lee, I.Y., Nair, V., Fischer, E., Boshoff, H.I.M., and Barry, C.E. (2012). Meropenem inhibits D,D-carboxypeptidase activity in Mycobacterium tuberculosis. Mol. Microbiol. *86*, 367–381.

Kwan, A.H., Mobli, M., Gooley, P.R., King, G.F., and Mackay, J.P. (2011). Macromolecular NMR spectroscopy for the non-spectroscopist: Macromolecular NMR for the non-spectroscopists I. FEBS J. *278*, 687–703.

Lairson, L.L., Henrissat, B., Davies, G.J., and Withers, S.G. (2008). Glycosyltransferases: Structures, Functions, and Mechanisms. Annu. Rev. Biochem. *77*, 521–555.

Landau, M., Mayrose, I., Rosenberg, Y., Glaser, F., Martz, E., Pupko, T., and Ben-Tal, N. (2005). ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. Nucleic Acids Res. *33*, W299–W302.

Lang, R. (2013). Recognition of the mycobacterial cord factor by Mincle: relevance for granuloma formation and resistance to tuberculosis. Front. Immunol. *4*, 5.

Lapp, D., Patterson, B.W., and Elbein, A.D. (1971). Properties of a Trehalose Phosphate Synthetase from Mycobacterium smegmatis ACTIVATION OF THE ENZYME BY POLYNUCLEOTIDES AND OTHER POLYANIONS. J. Biol. Chem. *246*, 4567–4579.

Lechartier, B., Hartkoorn, R.C., and Cole, S.T. (2012). In vitro combination studies of benzothiazinone lead compound BTZ043 against Mycobacterium tuberculosis. Antimicrob. Agents Chemother. *56*, 5790–5793.

Lechevalier, M.P., and Lechevalier, H. (1970). Chemical composition as a criterion in the classification of aerobic actinomycetes. Int. J. Syst. Evol. Microbiol. *20*, 435–443.

Lee, Y.C. (1966). Isolation and characterization of lipopolysaccharides containing 6-O-methyl-D-glucose from Mycobacterium species. J. Biol. Chem. *241*, 1899–1908.

Lee, Y.C., and Ballou, C.E. (1964). 6-O-METHYL-D-GLUCOSE FROM MYCOBACTERIA. J. Biol. Chem. *239*, PC3602-3603.

Lee, C.-D., Sun, H.-C., Hu, S.-M., Chiu, C.-F., Homhuan, A., Liang, S.-M., Leng, C.-H., and Wang, T.-F. (2008). An improved SUMO fusion protein system for effective production of native proteins. Protein Sci. *17*, 1241–1248.

Lee, R.E., Protopopova, M., Crooks, E., Slayden, R.A., Terrot, M., and Barry, C.E. (2003). Combinatorial lead optimization of [1,2]-diamines based on ethambutol as potential antituberculosis preclinical candidates. J. Comb. Chem. *5*, 172–187.

Lee, S.S., Hong, S.Y., Errey, J.C., Izumi, A., Davies, G.J., and Davis, B.G. (2011). Mechanistic evidence for a front-side, SNi-type reaction in a retaining glycosyltransferase. Nat. Chem. Biol. *7*, 631–638.

Lemassu, A., and Daffé, M. (1994). Structural features of the exocellular polysaccharides of Mycobacterium tuberculosis. Biochem. J. *297*, 351–357.

Lemassu, A., Lanéelle, M.A., and Daffé, M. (1991). Revised structure of a trehalose-containing immunoreactive glycolipid of Mycobacterium tuberculosis. FEMS Microbiol. Lett. *62*, 171–175.

Lew, J.M., Kapopoulou, A., Jones, L.M., and Cole, S.T. (2011). TubercuList--10 years after. Tuberc. Edinb. Scotl. *91*, 1–7.

Lodish, H. (2008). Molecular cell biology, 5th Ed. (Macmillan).

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. *42*, D490–D495.

Lovering, A.L., de Castro, L.H., Lim, D., and Strynadka, N.C.J. (2007). Structural Insight into the Transglycosylation Step of Bacterial Cell-Wall Biosynthesis. Science *315*, 1402–1405.

Luft, J.R., Wolfley, J.R., Said, M.I., Nagel, R.M., Lauricella, A.M., Smith, J.L., Thayer, M.H., Veatch, C.K., Snell, E.H., Malkowski, M.G., et al. (2007). Efficient optimization of crystallization conditions by manipulation of drop volume ratio and temperature. Protein Sci. *16*, 715–722.

Ma, Z., Lienhardt, C., McIlleron, H., Nunn, A.J., and Wang, X. (2010). Global tuberculosis drug development pipeline: the need and the reality. Lancet Lond. Engl. *375*, 2100–2109.

Machida, Y., and Bloch, K. (1973). Complex formation between mycobacterial polysaccharides and fatty acyl-CoA derivatives. Proc. Natl. Acad. Sci. *70*, 1146–1148.

Makarov, V., Manina, G., Mikusova, K., Möllmann, U., Ryabova, O., Saint-Joanis, B., Dhar, N., Pasca, M.R., Buroni, S., Lucarelli, A.P., et al. (2009). Benzothiazinones kill Mycobacterium tuberculosis by blocking arabinan synthesis. Science *324*, 801–804.

Manjunatha, U., Boshoff, H.I., and Barry, C.E. (2009). The mechanism of action of PA-824: Novel insights from transcriptional profiling. Commun. Integr. Biol. *2*, 215–218.

Manjunatha, U.H., Boshoff, H., Dowd, C.S., Zhang, L., Albert, T.J., Norton, J.E., Daniels, L., Dick, T., Pang, S.S., and Barry, C.E. (2006). Identification of a nitroimidazo-oxazine-specific protein involved in PA-824 resistance in Mycobacterium tuberculosis. Proc. Natl. Acad. Sci. U. S. A. *103*, 431–436.

Maranha, A., Moynihan, P.J., Miranda, V., Correia Lourenço, E., Nunes-Costa, D., Fraga, J.S., José Barbosa Pereira, P., Macedo-Ribeiro, S., Ventura, M.R., Clarke, A.J., et al. (2015). Octanoylation of early intermediates of mycobacterial methylglucose lipopolysaccharides. Sci. Rep. *5*, 13610.

Martin, J.L., and McMillan, F.M. (2002). SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. Curr. Opin. Struct. Biol. *12*, 783–793.

Maruta, K., Mitsuzumi, H., Nakada, T., Kubota, M., Chaen, H., Fukuda, S., Sugimoto, T., and Kurimoto, M. (1996c). Cloning and sequencing of a cluster of genes encoding novel enzymes of trehalose biosynthesis from thermophilic archaebacterium Sulfolobus acidocaldarius. Biochim. Biophys. Acta *1291*, 177–181.

Maruta, K., Hattori, K., Nakada, T., Kubota, M., Sugimoto, T., and Kurimoto, M. (1996a). Cloning and sequencing of trehalose biosynthesis genes from Arthrobacter sp. Q36. Biochim. Biophys. Acta *1289*, 10–13.

Maruta, K., Hattori, K., Nakada, T., Kubota, M., Sugimoto, T., and Kurimoto, M. (1996b). Cloning and sequencing of trehalose biosynthesis genes from Rhizobium sp. M-11. Biosci. Biotechnol. Biochem. *60*, 717–720.

Matsumoto, M., Hashizume, H., Tomishige, T., Kawasaki, M., Tsubouchi, H., Sasaki, H., Shimokawa, Y., and Komatsu, M. (2006). OPC-67683, a nitro-dihydro-imidazooxazole derivative with promising action against tuberculosis in vitro and in mice. PLoS Med. *3*, e466.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). *Phaser* crystallographic software. J. Appl. Crystallogr. *40*, 658–674.

McIlleron, H., Meintjes, G., Burman, W.J., and Maartens, G. (2007). Complications of antiretroviral therapy in patients with tuberculosis: drug interactions, toxicity, and immune reconstitution inflammatory syndrome. J. Infect. Dis. *196 Suppl 1*, S63-75.

McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y.M., Buso, N., Cowley, A.P., and Lopez, R. (2013). Analysis Tool Web Services from the EMBL-EBI. Nucleic Acids Res. *41*, W597–W600.

Mdluli, K., Kaneko, T., and Upton, A. (2015). The Tuberculosis Drug Discovery and Development Pipeline and Emerging Drug Targets. Cold Spring Harb. Perspect. Med. *5*, a021154–a021154.

Mendes, V., Maranha, A., Alarico, S., da Costa, M.S., and Empadinhas, N. (2011). Mycobacterium tuberculosis Rv2419c, the missing glucosyl-3-phosphoglycerate phosphatase for the second step in methylglucose lipopolysaccharide biosynthesis. Sci. Rep. *1*, 177.

Mendes, V., Maranha, A., Alarico, S., and Empadinhas, N. (2012). Biosynthesis of mycobacterial methylglucose lipopolysaccharides. Nat. Prod. Rep. *29*, 834.

Mertens, H.D.T., and Svergun, D.I. (2010). Structural characterization of proteins and complexes using small-angle X-ray solution scattering. J. Struct. Biol. *172*, 128–141.

Miah, F., Koliwer-Brandl, H., Rejzek, M., Field, R.A., Kalscheuer, R., and Bornemann, S. (2013). Flux through Trehalose Synthase Flows from Trehalose to the Alpha Anomer of Maltose in Mycobacteria. Chem. Biol. *20*, 487–493.

Migliardo, F., Salmeron, C., and Bayan, N. (2015). Mobility and temperature resistance of trehalose mycolates as key characteristics of the outer membrane of Mycobacterium tuberculosis. J. Biomol. Struct. Dyn. *33*, 447–459.

Milev, S. (2013). Isothermal titration calorimetry: Principles and experimental design. Gen. Electr. *9*.

Moréra, S., Larivière, L., Kurzeck, J., Aschke-Sonnenborn, U., Freemont, P.S., Janin, J., and Rüger, W. (2001). High resolution crystal structures of T4 phage β-glucosyltransferase: induced fit and effect of substrate and metal binding. J. Mol. Biol. *311*, 569–577.

Muñoz, M., Lanéelle, M.A., Luquin, M., Torrelles, J., Julián, E., Ausina, V., and Daffé, M. (1997). Occurrence of an antigenic triacyl trehalose in clinical isolates and reference strains of Mycobacterium tuberculosis. FEMS Microbiol. Lett. *157*, 251–259.

Murphy, T.A., and Wyatt, G.R. (1965). THE ENZYMES OF GLYCOGEN AND TREHALOSE SYNTHESIS IN SILK MOTH FAT BODY. J. Biol. Chem. *240*, 1500–1508.

Murphy, H.N., Stewart, G.R., Mischenko, V.V., Apt, A.S., Harris, R., McAlister, M.S.B., Driscoll, P.C., Young, D.B., and Robertson, B.D. (2005). The OtsAB Pathway Is Essential for Trehalose Biosynthesis in Mycobacterium tuberculosis. J. Biol. Chem. *280*, 14524–14529.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D Biol. Crystallogr. *53*, 240–255.

Nannelli, A., Chirulli, V., Longo, V., and Gervasi, P.G. (2008). Expression and induction by rifampicin of CAR- and PXR-regulated CYP2B and CYP3A in liver, kidney and airways of pig. Toxicology *252*, 105–112.

Nannenga, B.L., and Gonen, T. (2014). Protein structure determination by MicroED. Curr. Opin. Struct. Biol. *27*, 24–31.

Neres, J., Pojer, F., Molteni, E., Chiarelli, L.R., Dhar, N., Boy-Röttger, S., Buroni, S., Fullam, E., Degiacomi, G., Lucarelli, A.P., et al. (2012). Structural basis for benzothiazinone-mediated killing of Mycobacterium tuberculosis. Sci. Transl. Med. *4*, 150ra121.

Niesen, F.H., Berglund, H., and Vedadi, M. (2007). The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nat. Protoc. *2*, 2212–2221.

Nobre, A., Alarico, S., Maranha, A., Mendes, V., and Empadinhas, N. (2014). The molecular biology of mycobacterial trehalose in the quest for advanced tuberculosis therapies. Microbiology *160*, 1547–1570.

Nuermberger, E.L., Yoshimatsu, T., Tyagi, S., O'Brien, R.J., Vernon, A.N., Chaisson, R.E., Bishai, W.R., and Grosset, J.H. (2004). Moxifloxacin-containing regimen greatly reduces time to culture conversion in murine tuberculosis. Am. J. Respir. Crit. Care Med. *169*, 421–426.

Ogata, H., Nishikawa, K., and Lubitz, W. (2015). Hydrogens detected by subatomic resolution protein crystallography in a [NiFe] hydrogenase. Nature *520*, 571–574.

Ohtake, S., and Wang, Y.J. (2011). Trehalose: Current Use and Future Applications. J. Pharm. Sci. *100*, 2020–2053.

Ortalo-Magne, A., Dupont, M.-A., Lemassu, A., Andersen, A.B., Gounon, P., and Mamadou, D. (1995). Molecular composition of the outermost capsular material of the tubercle bacillus. Microbiology *141*, 1609–1620.

Pabst, M.J., Gross, J.M., Brozna, J.P., and Goren, M.B. (1988). Inhibition of macrophage priming by sulfatide from Mycobacterium tuberculosis. J. Immunol. Baltim. Md 1950 *140*, 634–640.

Palomino, J.C., Leão, S.C., and Ritacco, V. (2007). Tuberculosis 2007; from basic science to patient care.

Pan, Y.T., Drake, R.R., and Elbein, A.D. (1996). Trehalose-P synthase of mycobacteria: its substrate specificity is affected by polyanions. Glycobiology *6*, 453–461.

Pan, Y.T., Carroll, J.D., and Elbein, A.D. (2002). Trehalose-phosphate synthase of Mycobacterium tuberculosis: Cloning, expression and properties of the recombinant enzyme. Eur. J. Biochem. *269*, 6091–6100.

Pan, Y.T., Koroth Edavana, V., Jourdian, W.J., Edmondson, R., Carroll, J.D., Pastuszak, I., and Elbein, A.D. (2004). Trehalose synthase of Mycobacterium smegmatis: purification, cloning, expression, and properties of the enzyme. Eur. J. Biochem. FEBS *271*, 4259–4269.

Pan, Y.-T., Carroll, J.D., Asano, N., Pastuszak, I., Edavana, V.K., and Elbein, A.D. (2008). Trehalose synthase converts glycogen to trehalose: TreS converts glycogen to trehalose. FEBS J. *275*, 3408–3420.

Parish, T., Liu, J., Nikaido, H., and Stoker, N.G. (1997). A Mycobacterium smegmatis mutant with a defective inositol monophosphate phosphatase gene homolog has altered cell envelope permeability. J. Bacteriol. *179*, 7827–7833.

Parveen, N., and Cornell, K.A. (2011). Methylthioadenosine/S-adenosylhomocysteine nucleosidase, a critical enzyme for bacterial metabolism: Involvement of MTA/SAH nucleosidase in bacterial metabolism. Mol. Microbiol. *79*, 7–20.

Patterson, J.H., Waller, R.F., Jeevarajah, D., Billman-Jacobe, H., and McConville, M.J. (2003). Mannose metabolism is required for mycobacterial growth. Biochem. J. *372*, 77–86.

Peränen, J., Rikkonen, M., Hyvönen, M., and Kääriäinen, L. (1996). T7 Vectors with a Modified T7lacPromoter for Expression of Proteins inEscherichia coli. Anal. Biochem. *236*, 371–373.

Petoukhov, M.V., and Svergun, D.I. (2013). Applications of small-angle X-ray scattering to biomacromolecular solutions. Int. J. Biochem. Cell Biol. *45*, 429–437.

Petoukhov, M.V., Franke, D., Shkumatov, A.V., Tria, G., Kikhney, A.G., Gajda, M., Gorba, C., Mertens, H.D.T., Konarev, P.V., and Svergun, D.I. (2012). New developments in the *ATSAS* program package for small-angle scattering data analysis. J. Appl. Crystallogr. *45*, 342–350.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera-A visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612.

Petzold, E.W., Himmelreich, U., Mylonakis, E., Rude, T., Toffaletti, D., Cox, G.M., Miller, J.L., and Perfect, J.R. (2006). Characterization and Regulation of the Trehalose Synthesis Pathway and Its Importance in the Pathogenicity of Cryptococcus neoformans. Infect. Immun. *74*, 5877–5887.

Piccaro, G., Giannoni, F., Filippini, P., Mustazzolu, A., and Fattorini, L. (2013). Activities of drug combinations against Mycobacterium tuberculosis grown in aerobic and hypoxic acidic conditions. Antimicrob. Agents Chemother. *57*, 1428–1433.

Pierce, M.M., Raman, C.S., and Nall, B.T. (1999). Isothermal titration calorimetry of protein-protein interactions. Methods San Diego Calif *19*, 213–221.

Pitarque, S., Larrouy-Maumus, G., Payré, B., Jackson, M., Puzo, G., and Nigou, J. (2008). The immunomodulatory lipoglycans, lipoarabinomannan and lipomannan, are exposed at the mycobacterial cell surface. Tuberc. Edinb. Scotl. *88*, 560–565.

Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., et al. (2012). The Pfam protein families database. Nucleic Acids Res. *40*, D290-301.

Pym, A.S., Diacon, A.H., Tang, S.-J., Conradie, F., Danilovits, M., Chuchottaworn, C., Vasilyeva, I., Andries, K., Bakare, N., De Marez, T., et al. (2016). Bedaquiline in the treatment of multidrug- and extensively drug-resistant tuberculosis. Eur. Respir. J. *47*, 564–574.

Rastogi, N., Legrand, E., and Sola, C. (2001). The mycobacteria: an introduction to nomenclature and pathogenesis. Rev. Sci. Tech. Int. Off. Epizoot. *20*, 21–54.

Ren, H., Dover, L.G., Islam, S.T., Alexander, D.C., Chen, J.M., Besra, G.S., and Liu, J. (2007). Identification of the lipooligosaccharide biosynthetic gene cluster from Mycobacterium marinum. Mol. Microbiol. *63*, 1345–1359.

Rengarajan, J., Bloom, B.R., and Rubin, E.J. (2005). Genome-wide requirements for Mycobacterium tuberculosis adaptation and survival in macrophages. Proc. Natl. Acad. Sci. U. S. A. *102*, 8327–8332.

Rhodes, G. (2006). Crystallography Made Crystal Clear: A Guide for Users of Macromolecular Models, 3rd Ed. (Academic Press).

Riccardi, G., Pasca, M.R., and Buroni, S. (2009). Mycobacterium tuberculosis: drug resistance and future perspectives. Future Microbiol. *4*, 597–614.

Riley, L.W. (2006). Of mice, men, and elephants: Mycobacterium tuberculosis cell envelope lipids and pathogenesis. J. Clin. Invest. *116*, 1475–1478.

Rombouts, Y., Burguière, A., Maes, E., Coddeville, B., Elass, E., Guérardel, Y., and Kremer, L. (2009). Mycobacterium marinum lipooligosaccharides are unique caryophyllose-containing cell wall glycolipids that inhibit tumor necrosis factor-alpha secretion in macrophages. J. Biol. Chem. *284*, 20975–20988.

Rombouts, Y., Elass, E., Biot, C., Maes, E., Coddeville, B., Burguière, A., Tokarski, C., Buisine, E., Trivelli, X., Kremer, L., et al. (2010). Structural analysis of an unusual bioactive N-acylated lipo-oligosaccharide LOS-IV in Mycobacterium marinum. J. Am. Chem. Soc. *132*, 16073–16084.

Rombouts, Y., Alibaud, L., Carrère-Kremer, S., Maes, E., Tokarski, C., Elass, E., Kremer, L., and Guérardel, Y. (2011). Fatty acyl chains of Mycobacterium marinum lipooligosaccharides: structure, localization and acylation by PapA4 (MMAR_2343) protein. J. Biol. Chem. *286*, 33678–33688.

Rose, P.W., Bi, C., Bluhm, W.F., Christie, C.H., Dimitropoulos, D., Dutta, S., Green, R.K., Goodsell, D.S., Prlic, A., Quesada, M., et al. (2013). The RCSB Protein Data Bank: new resources for research and education. Nucleic Acids Res. *41*, D475–D482.

Rosenthal, I.M., Zhang, M., Williams, K.N., Peloquin, C.A., Tyagi, S., Vernon, A.A., Bishai, W.R., Chaisson, R.E., Grosset, J.H., and Nuermberger, E.L. (2007). Daily dosing of rifapentine cures tuberculosis in three months or less in the murine model. PLoS Med. *4*, e344.

Rosenzweig, S.D., Yancoski, J., Bernasconi, A., Krasovec, S., Marciano, B.E., Casimir, L., Berberian, G., Símboli, N., Rousseau, M., and Calle, G. (2006). Thirteen years of culture-positive M. bovis-BCG infection in an IL-12Rbeta1 deficient patient: treatment and outcome. J. Infect. *52*, e69-72.

Roth, R., and Sussman, M. (1966). Trehalose synthesis in the cellular slime mold Dictyostelium discoideum. Biochim. Biophys. Acta *122*, 225–231.

Rousseau, C., Neyrolles, O., Bordat, Y., Giroux, S., Sirakova, T.D., Prevost, M.-C., Kolattukudy, P.E., Gicquel, B., and Jackson, M. (2003). Deficiency in mycolipenate-and mycosanoate-derived acyltrehaloses enhances early interactions of Mycobacterium tuberculosis with host cells. Cell. Microbiol. *5*, 405–415.

Rozwarski, D.A., Vilchèze, C., Sugantino, M., Bittman, R., and Sacchettini, J.C. (1999). Crystal structure of the Mycobacterium tuberculosis enoyl-ACP reductase, InhA, in complex with NAD+ and a C16 fatty acyl substrate. J. Biol. Chem. *274*, 15582–15589.

Rupp, B. (2009). Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology (Garland Science).

Ruusala, T., and Kurland, C.G. (1984). Streptomycin preferentially perturbs ribosomal proofreading. Mol. Gen. Genet. MGG *198*, 100–104.

S. Glickman, M. (2008). 5 Cording, Cord Factors, and Trehalose Dimycolate. In The Mycobacterial Cell Envelope, G. Avenir, M. Daffé, and J.-M. Reyrat, eds. (American Society of Microbiology), pp. 63–73.

Saavedra, R., Segura, E., Leyva, R., Esparza, L.A., and López-Marín, L.M. (2001). Mycobacterial di-O-acyl-trehalose inhibits mitogen- and antigen-induced proliferation of murine T cells in vitro. Clin. Diagn. Lab. Immunol. *8*, 1081–1088.

Sakamoto, K., Kim, M.J., Rhoades, E.R., Allavena, R.E., Ehrt, S., Wainwright, H.C., Russell, D.G., and Rohde, K.H. (2013). Mycobacterial trehalose dimycolate reprograms macrophage global gene expression and activates matrix metalloproteinases. Infect. Immun. *81*, 764–776.

Sali, A., and Blundell, T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. *234*, 779–815.

Sambou, T., Dinadayala, P., Stadthagen, G., Barilone, N., Bordat, Y., Constant, P., Levillain, F., Neyrolles, O., Gicquel, B., Lemassu, A., et al. (2008). Capsular glucan and intracellular glycogen of Mycobacterium tuberculosis: biosynthesis and impact on the persistence in mice. Mol. Microbiol. *70*, 762–774.

Sani, M., Houben, E.N.G., Geurtsen, J., Pierson, J., de Punder, K., van Zon, M., Wever, B., Piersma, S.R., Jiménez, C.R., Daffé, M., et al. (2010). Direct Visualization by Cryo-EM of the Mycobacterial Capsular Layer: A Labile Structure Containing ESX-1-Secreted Proteins. PLoS Pathog. *6*, e1000794.

Santivañez-Veliz, M., Pérez-Silanes, S., Torres, E., and Moreno-Viguri, E. (2016). Design and synthesis of novel quinoxaline derivatives as potential candidates for treatment of multidrug-resistant and latent tuberculosis. Bioorg. Med. Chem. Lett. *26*, 2188–2193.

Sassetti, C.M., Boyd, D.H., and Rubin, E.J. (2003). Genes required for mycobacterial growth defined by high density mutagenesis. Mol. Microbiol. *48*, 77–84.

Schubert, H.L., Blumenthal, R.M., and Cheng, X. (2003). Many paths to methyltransfer: a chronicle of convergence. Trends Biochem. Sci. *28*, 329–335.

Schuck, P. (2000). Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. Biophys. J. *78*, 1606–1619.

Schuck, P. (2016). Sedimentation coefficient distributions of large particles. The Analyst *141*, 4400–4409.

Scorpio, A., and Zhang, Y. (1996). Mutations in pncA, a gene encoding pyrazinamidase/nicotinamidase, cause resistance to the antituberculous drug pyrazinamide in tubercle bacillus. Nat. Med. *2*, 662–667.

Scorpio, A., Lindholm-Levy, P., Heifets, L., Gilman, R., Siddiqi, S., Cynamon, M., and Zhang, Y. (1997). Characterization of pncA mutations in pyrazinamide-resistant Mycobacterium tuberculosis. Antimicrob. Agents Chemother. *41*, 540–543.

Scott, D.E., Coyne, A.G., Hudson, S.A., and Abell, C. (2012). Fragment-Based Approaches in Drug Discovery and Chemical Biology. Biochemistry (Mosc.) *51*, 4990–5003.

Seibold, G.M., and Eikmanns, B.J. (2007). The glgX gene product of Corynebacterium glutamicum is required for glycogen degradation and for fast adaptation to hyperosmotic stress. Microbiol. Read. Engl. *153*, 2212–2220.

Shi, J., Blundell, T.L., and Mizuguchi, K. (2001). FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties11Edited by B. Honig. J. Mol. Biol. *310*, 243–257.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., et al. (2014). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. *7*, 539–539.

Silva, J.P., Appelberg, R., and Gama, F.M. (2016). Antimicrobial peptides as novel anti-tuberculosis therapeutics. Biotechnol. Adv.

Silva, Z., Alarico, S., and da Costa, M.S. (2005). Trehalose biosynthesis in Thermus thermophilus RQ-1: biochemical properties of the trehalose-6-phosphate synthase and trehalose-6-phosphate phosphatase. Extrem. Life Extreme Cond. *9*, 29–36.

Singh, R., Manjunatha, U., Boshoff, H.I.M., Ha, Y.H., Niyomrattanakit, P., Ledwidge, R., Dowd, C.S., Lee, I.Y., Kim, P., Zhang, L., et al. (2008). PA-824 kills nonreplicating Mycobacterium tuberculosis by intracellular NO release. Science *322*, 1392–1395.

Sirgel, F.A., Tait, M., Warren, R.M., Streicher, E.M., Böttger, E.C., van Helden, P.D., Gey van Pittius, N.C., Coetzee, G., Hoosain, E.Y., Chabula-Nxiweni, M., et al. (2012). Mutations in the rrs A1401G gene and phenotypic resistance to amikacin and capreomycin in Mycobacterium tuberculosis. Microb. Drug Resist. Larchmt. N *18*, 193–197.

Skripconoka, V., Danilovits, M., Pehme, L., Tomson, T., Skenders, G., Kummik, T., Cirule, A., Leimane, V., Kurve, A., Levina, K., et al. (2013). Delamanid improves outcomes and reduces mortality in multidrug-resistant tuberculosis. Eur. Respir. J. *41*, 1393–1400.

Sloan, D.J., Davies, G.R., and Khoo, S.H. (2013). Recent advances in tuberculosis: New drugs and treatment regimens. Curr. Respir. Med. Rev. *9*, 200–210.

Smith, W.L., and Ballou, C.E. (1973). The 6-O-methylglucose-containing lipopolysaccharides of Mycobacterium phlei. Locations of the neutral and acidic acyl groups. J. Biol. Chem. *248*, 7118–7125.

Sotgiu, G., Centis, R., D'Ambrosio, L., Alffenaar, J.-W.C., Anger, H.A., Caminero, J.A., Castiglia, P., De Lorenzo, S., Ferrara, G., Koh, W.-J., et al. (2012). Efficacy, safety and tolerability of linezolid containing regimens in treating MDR-TB and XDR-TB: systematic review and meta-analysis. Eur. Respir. J. *40*, 1430–1442.

Stadthagen, G., Sambou, T., Guerin, M., Barilone, N., Boudou, F., Kordulakova, J., Charles, P., Alzari, P.M., Lemassu, A., Daffe, M., et al. (2007). Genetic Basis for the Biosynthesis of Methylglucose Lipopolysaccharides in Mycobacterium tuberculosis. J. Biol. Chem. *282*, 27270–27276.

Stanley, S.A., and Cox, J.S. (2013). Host-pathogen interactions during Mycobacterium tuberculosis infections. Curr. Top. Microbiol. Immunol. *374*, 211–241.

Stehr, M., A Elamin, A., and Singh, M. (2014). Filling the pipeline-new drugs for an old disease. Curr. Top. Med. Chem. *14*, 110–129.

Steingart, K.R., Jotblad, S., Robsky, K., Deck, D., Hopewell, P.C., Huang, D., and Nahid, P. (2011). Higher-dose rifampin for the treatment of pulmonary tuberculosis: a systematic review. Int. J. Tuberc. Lung Dis. Off. J. Int. Union Tuberc. Lung Dis. *15*, 305–316.

Sterling, T.R., Villarino, M.E., Borisov, A.S., Shang, N., Gordin, F., Bliven-Sizemore, E., Hackman, J., Hamilton, C.D., Menzies, D., Kerrigan, A., et al. (2011). Three months of rifapentine and isoniazid for latent tuberculosis infection. N. Engl. J. Med. *365*, 2155–2166.

Stratton, H.M., Brooks, P.R., Griffiths, P.C., and Seviour, R.J. (2002). Cell surface hydrophobicity and mycolic acid composition of Rhodococcus strains isolated from activated sludge foam. J. Ind. Microbiol. Biotechnol. *28*, 264–267.

Strom, A.R., and Kaasen, I. (1993). Trehalose metabolism in Escherichia coli: stress protection and stress regulation of gene expression. Mol. Microbiol. *8*, 205–210.

Stuurman, A.L., Vonk Noordegraaf-Schouten, M., van Kessel, F., Oordt-Speets, A.M., Sandgren, A., and van der Werf, M.J. (2016). Interventions for improving adherence to treatment for latent tuberculosis infection: a systematic review. BMC Infect. Dis. *16*.

Sugahara, M. (2014). A Technique for High-Throughput Protein Crystallization in Ionically Cross-Linked Polysaccharide Gel Beads for X-Ray Diffraction Experiments. PLoS ONE *9*, e95017.

Sun, Z., Scorpio, A., and Zhang, Y. (1997). The pncA gene from naturally pyrazinamide-resistant Mycobacterium avium encodes pyrazinamidase and confers pyrazinamide susceptibility to resistant M. tuberculosis complex organisms. Microbiol. Read. Engl. *143 (Pt 10)*, 3367–3373.

Svergun, D.I. (1999). Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. Biophys. J. *76*, 2879–2886.

Svergun, D.I., and Koch, M.H.J. (2003). Small-angle scattering studies of biological macromolecules in solution. Rep. Prog. Phys. *66*, 1735.

Tahlan, K., Wilson, R., Kastrinsky, D.B., Arora, K., Nair, V., Fischer, E., Barnes, S.W., Walker, J.R., Alland, D., Barry, C.E., et al. (2012). SQ109 targets MmpL3, a membrane transporter of trehalose monomycolate involved in mycolic acid donation to the cell wall core of Mycobacterium tuberculosis. Antimicrob. Agents Chemother. *56*, 1797–1809.

Tengel, T. (2007). Studies of protein structure, dynamics and protein-ligand interactions using NMR spectroscopy. Univ.

Toriyama, S., Yano, I., Masui, M., Kusunose, E., Kusunose, M., and Akimori, N. (1980). Regulation of cell wall mycolic acid biosynthesis in acid-fast bacteria. I. Temperature-induced changes in mycolic acid molecular species and related compounds in Mycobacterium phlei. J. Biochem. (Tokyo) *88*, 211–221.

Torrelles, J.B., Knaup, R., Kolareth, A., Slepushkina, T., Kaufman, T.M., Kang, P., Hill, P.J., Brennan, P.J., Chatterjee, D., Belisle, J.T., et al. (2008). Identification of Mycobacterium tuberculosis clinical isolates with altered phagocytosis by human macrophages due to a truncated lipoarabinomannan. J. Biol. Chem. *283*, 31417–31428.

Trefzer, C., Škovierová, H., Buroni, S., Bobovská, A., Nenci, S., Molteni, E., Pojer, F., Pasca, M.R., Makarov, V., Cole, S.T., et al. (2012). Benzothiazinones are suicide inhibitors of mycobacterial decaprenylphosphoryl-β-D-ribofuranose 2'-oxidase DprE1. J. Am. Chem. Soc. *134*, 912–915.

Tuffal, G., Ponthus, C., Picard, C., Rivière, M., and Puzo, G. (1995). Structural elucidation of novel methylglucose-containing polysaccharides from Mycobacterium xenopi. Eur. J. Biochem. FEBS *233*, 377–383.

Tung, K.K., and Ballou, C.E. (1973). Biosynthesis of a mycobacterial lipopolysaccharide. Properties of the polysaccharide: acyl coenzyme A acyltransferase reaction. J. Biol. Chem. *248*, 7126–7133.

Tzvetkov, M., Klopprogge, C., Zelder, O., and Liebl, W. (2003). Genetic dissection of trehalose biosynthesis in Corynebacterium glutamicum: inactivation of trehalose production leads to impaired growth and an altered cell wall lipid composition. Microbiol. Read. Engl. *149*, 1659–1673.

Uhlin, M., Andersson, J., Zumla, A., and Maeurer, M. (2012). Adjunct Immunotherapies for Tuberculosis. J. Infect. Dis. *205*, S325–S334.

Varki, A., Cummings, R., Esko, J., Freeze, H., Stanley, P., Bertozzi, C.R., Hart, G., and Etzler, M.E. (2009). Essentials of Glycobiology, 2nd Ed (Cold Spring Harbor Laboratory Press).

Verma, R.K., Germishuizen, W.A., Motheo, M.P., Agrawal, A.K., Singh, A.K., Mohan, M., Gupta, P., Gupta, U.D., Cholo, M., Anderson, R., et al. (2013). Inhaled microparticles containing clofazimine are efficacious in treatment of experimental tuberculosis in mice. Antimicrob. Agents Chemother. *57*, 1050–1052.

Vogel, G., Fiehn, O., Jean-Richard-dit-Bressel, L., Boller, T., Wiemken, A., Aeschbacher, R.A., and Wingler, A. (2001). Trehalose metabolism in Arabidopsis: occurrence of trehalose and molecular cloning and characterization of trehalose-6-phosphate synthase homologues. J. Exp. Bot. *52*, 1817–1826.

Voskuil, M.I., Schnappinger, D., Visconti, K.C., Harrell, M.I., Dolganov, G.M., Sherman, D.R., and Schoolnik, G.K. (2003). Inhibition of Respiration by Nitric Oxide Induces a *Mycobacterium tuberculosis* Dormancy Program. J. Exp. Med. *198*, 705–713.

Wallis, R.S., and Hafner, R. (2015). Advancing host-directed therapy for tuberculosis. Nat. Rev. Immunol. *15*, 255–263.

Wallis, R.S., Jakubiec, W.M., Kumar, V., Silvia, A.M., Paige, D., Dimitrova, D., Li, X., Ladutko, L., Campbell, S., Friedland, G., et al. (2010). Pharmacokinetics and whole-blood bactericidal activity against Mycobacterium tuberculosis of single doses of PNU-100480 in healthy volunteers. J. Infect. Dis. *202*, 745–751.

Wallis, R.S., Jakubiec, W., Kumar, V., Bedarida, G., Silvia, A., Paige, D., Zhu, T., Mitton-Fry, M., Ladutko, L., Campbell, S., et al. (2011). Biomarker-assisted dose selection for safety and efficacy in early development of PNU-100480 for tuberculosis. Antimicrob. Agents Chemother. *55*, 567–574.

Wallis, R.S., Dawson, R., Friedrich, S.O., Venter, A., Paige, D., Zhu, T., Silvia, A., Gobey, J., Ellery, C., Zhang, Y., et al. (2014). Mycobactericidal activity of sutezolid (PNU-100480) in sputum (EBA) and blood (WBA) of patients with pulmonary tuberculosis. PloS One *9*, e94462.

Wallis, R.S., Maeurer, M., Mwaba, P., Chakaya, J., Rustomjee, R., Migliori, G.B., Marais, B., Schito, M., Churchyard, G., Swaminathan, S., et al. (2016). Tuberculosis—advances in development of new drugs, treatment regimens, host-directed therapies, and biomarkers. Lancet Infect. Dis. *16*, e34–e46.

Wang, F., Langley, R., Gulten, G., Dover, L.G., Besra, G.S., Jacobs, W.R., and Sacchettini, J.C. (2007). Mechanism of thioamide drug action against tuberculosis and leprosy. J. Exp. Med. *204*, 73–78.

Ward, J.J., Sodhi, J.S., McGuffin, L.J., Buxton, B.F., and Jones, D.T. (2004). Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. J. Mol. Biol. *337*, 635–645.

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189–1191.

Węgrzyn, A., Szalewska-Pałasz, A., Błaszczak, A., Liberek, K., and Węgrzyn, G. (1998). Differential inhibition of transcription from σ70- and σ32-dependent promoters by rifampicin. FEBS Lett. *440*, 172–174.

Weisman, L.S., and Ballou, C.E. (1984). Biosynthesis of the mycobacterial methylmannose polysaccharide. Identification of an alpha 1----4-mannosyltransferase. J. Biol. Chem. *259*, 3457–3463.

Welsh, K.J., Hunter, R.L., and Actor, J.K. (2013). Trehalose 6,6'-dimycolate--a coat to regulate tuberculosis immunopathogenesis. Tuberc. Edinb. Scotl. *93 Suppl*, S3-9.

Whitmore, L., and Wallace, B.A. (2004). DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. Nucleic Acids Res. *32*, W668–W673.

Whitmore, L., and Wallace, B.A. (2008). Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases. Biopolymers *89*, 392–400.

Williamson, M.P. (2013). Using chemical shift perturbation to characterise ligand binding. Prog. Nucl. Magn. Reson. Spectrosc. *73*, 1–16.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., et al. (2011). Overview of the *CCP* 4 suite and current developments. Acta Crystallogr. D Biol. Crystallogr. *67*, 235–242.

Woodruff, P.J., Carlson, B.L., Siridechadilok, B., Pratt, M.R., Senaratne, R.H., Mougous, J.D., Riley, L.W., Williams, S.J., and Bertozzi, C.R. (2004). Trehalose Is Required for Growth of Mycobacterium smegmatis. J. Biol. Chem. *279*, 28835–28843.

Wright, B.E., and Marshall, R. (1971). Trehalose Synthesis during Differentiation in Dictyostelium discoideum I. ANALYSIS AND PREDICTIONS BY COMPUTER SIMULATION. J. Biol. Chem. *246*, 5335–5339.

Wüthrich, K. (1986). NMR of Proteins and Nucleic Acids (Wiley).

Wüthrich, K. (2001). Nuclear Magnetic Resonance (NMR) Spectroscopy of Proteins. eLS.

Yabusaki, K.K., and Ballou, C.E. (1979). Effect of polymethylpolysaccharides on the hydrolysis of palmitoyl coenzyme A by a thioesterase from Mycobacterium smegmatis. J. Biol. Chem. *254*, 12314–12317.

Yabusaki, K.K., Cohen, R.E., and Ballou, C.E. (1979). Conformational changes associated with complex formation between a mycobacterial polymethylpolysaccharide and palmitic acid. J. Biol. Chem. *254*, 7282–7286.

Young, D. (2009). Animal models of tuberculosis. Eur. J. Immunol. *39*, 2011–2014.

Yuan, Y., Barrett, D., Zhang, Y., Kahne, D., Sliz, P., and Walker, S. (2007). Crystal structure of a peptidoglycan glycosyltransferase suggests a model for processive glycan chain synthesis. Proc. Natl. Acad. Sci. *104*, 5348–5353.

Zhang, L., Goren, M.B., Holzer, T.J., and Andersen, B.R. (1988). Effect of Mycobacterium tuberculosis-derived sulfolipid I on human phagocytic cells. Infect. Immun. *56*, 2876–2883.

Zhang, M., Yang, Y., Xu, Y., Qie, Y., Wang, J., Zhu, B., Wang, Q., Jin, R., Xu, S., and Wang, H. (2007). Trehalose-6-phosphate Phosphatase from *Mycobacterium tuberculosis* induces humoral and cellular immune responses. FEMS Immunol. Med. Microbiol. *49*, 68–74.

Zheng, L., Zhou, X., Zhang, H., Ji, X., Li, L., Huang, L., Bai, L., and Zhang, H. (2012). Structural and Functional Analysis of Validoxylamine A 7′-phosphate Synthase ValL Involved in Validamycin A Biosynthesis. PLoS ONE *7*, e32033.

Zuber, B., Chami, M., Houssin, C., Dubochet, J., Griffiths, G., and Daffe, M. (2008). Direct Visualization of the Outer Membrane of Mycobacteria and Corynebacteria in Their Native State. J. Bacteriol. *190*, 5672–5680.

Zuiderweg, E.R.P. (2002). Mapping Protein−Protein Interactions in Solution by NMR Spectroscopy †. Biochemistry (Mosc.) *41*, 1–7.

Zumla, A., Abubakar, I., Raviglione, M., Hoelscher, M., Ditiu, L., McHugh, T.D., Squire, S.B., Cox, H., Ford, N., McNerney, R., et al. (2012). Drug-resistant tuberculosis--current dilemmas, unanswered questions, challenges, and priority needs. J. Infect. Dis. *205 Suppl 2*, S228-240.

Zumla, A., Nahid, P., and Cole, S.T. (2013). Advances in the development of new tuberculosis drugs and treatment regimens. Nat. Rev. Drug Discov. *12*, 388–404.

Zumla, A., Chakaya, J., Hoelscher, M., Ntoumi, F., Rustomjee, R., Vilaplana, C., Yeboah-Manu, D., Rasolof, V., Munderi, P., Singh, N., et al. (2015a). Towards host-directed therapies for tuberculosis. Nat. Rev. Drug Discov. *14*, 511–512.

Zumla, A., Maeurer, M., and Host-Directed Therapies Network (HDT-NET) Consortium (2015b). Host-Directed Therapies for Tackling Multi-Drug Resistant Tuberculosis: Learning From the Pasteur-Bechamp Debates. Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am. *61*, 1432–1438.

Zumla, A.I., Gillespie, S.H., Hoelscher, M., Philips, P.P., Cole, S.T., Abubakar, I., McHugh, T.D., Schito, M., Maeurer, M., and Nunn, A.J. (2014). New antituberculosis drugs, regimens, and adjunct therapies: needs, advances, and future prospects. Lancet Infect. Dis. *14*, 327–340.

Zuniga, E.S., Early, J., and Parish, T. (2015). The future for early-stage tuberculosis drug discovery. Future Microbiol. *10*, 217–229.

Zureck, A. (1985). G. P. Kubica and L. G. Wayne (Editors), The Mycobacteria — a Sourcebook, Part A and Part B (Microbiology Series Volume 15). 1553 S., 225 Abb., 122 Tab. New York-Basel 1984. Marcel Dekker. $ 295.00. ISBN: 0-8247-7009-9 (Part A), 0-8247-1917-4 (Part B). J. Basic Microbiol. *25*, 662–662.